FIELDS *and* WAVES *in* MODERN RADIO

GENERAL ELECTRIC SERIES

FIELDS *and* WAVES *in* MODERN RADIO

BY

SIMON RAMO
Director of Research,
Radio Division,
Hughes Aircraft Company

AND

JOHN R. WHINNERY
Associate Professor of Electrical Engineering
University of California

One of a series written in the interest of the General Electric Advanced Engineering Program

NEW YORK
JOHN WILEY AND SONS, INC.
LONDON: CHAPMAN AND HALL, LIMITED

SIXTH PRINTING, SEPTEMBER, 1949

PRINTED IN THE UNITED STATES OF AMERICA

PREFACE

This book is concerned with certain aspects of electromagnetic theory in their relation to the problems of modern radio and electronics engineering.

Several years ago, when this book was begun, there was no book which even approximately filled the very apparent need for an extensive treatment of field and wave theory from the radio engineer's point of view; all the needed information was distributed widely throughout the literature. Since that time several excellent textbooks have appeared, each of which has filled a part of that need, yet in the present volume the purpose, order of presentation, choice of material, and emphasis are different enough from those of the others so that it is not a duplication of these.

The present material was first prepared for use in company courses in which the students were at the same time starting their practice as engineers. The authors also had occasion to use essentially all the material of this book in their engineering analysis and laboratory work, in addition to trying to understand it from the slightly different point of view necessary for presentation in classes to other student engineers. The purpose was consequently one in which all the material should eventually be useful, either for specific design calculations or conceptually for a better general understanding and extension of techniques first learned empirically. Physical understanding was therefore more important than mathematical rigor, yet the mathematical treatments had to be specific enough for usefulness in the required quantitative calculations. Analogies to any of the well-grounded tools, techniques, or concepts of the engineer were to be made use of whenever it seemed that these could ease the way into a new and difficult subject.

Specifically, the most important objectives were the treatments of high-frequency circuits, skin effect, and shielding problems, problems of wave transmission and reflection, transmission lines and wave guides, cavity resonators, and antennas and other radiating systems, given in the latter two-thirds of this book; also — and this is important — to correlate fields and waves with circuits so that they are all seen as parts of a consistent whole. For study of these it is basically necessary only to present Maxwell's equations as the set of laws which apply, and then get on with the job. However, sad experience has made the authors believe that unless the preliminary introduction to the field equations is more extensive than a mere presentation, all further studies based upon

them are to an extent unsatisfying and insecure. Early chapters are consequently devoted to the less interesting job of presenting the basic laws in some detail, raising and answering some of the questions that the student will otherwise inevitably raise for himself.

This textbook is designed for students who have had the usual engineering mathematics courses through the calculus, but not necessarily any additional subjects such as vector analysis or extensive courses in differential equations. The required additional mathematics and vector language are woven into the presentations of physical laws and procedure in this text as much as seemed practicable, so that the strong complementary relations between the physics and mathematics might be made evident.

The first chapter is introduced to bridge certain gaps; in mathematics, between basic calculus and some of the mathematical tools required for the study of field and wave theory; in concept, between straightforward application of Kirchhoff's laws and the approximations and defined quantities of high-frequency circuits and transmission line problems. Elementary differential equation solutions, the Fourier series, and the use of complex exponentials are thus introduced with circuit and transmission line problems as oscillation and wave examples. The purpose is not, however, to present in completeness all important information on radio circuits and transmission lines. Since the material in the first chapter was designed for an average engineer or student who has heard of many of these items but is not completely prepared on some of them, it may be boring to one well acquainted with the techniques treated and should therefore be ignored or only skimmed; for one not at all familiar with any of the material, the objectives may be found to be too much for one chapter, and it should be supplemented by other textbooks. An engineer primarily interested in the high-frequency applications may also deal more lightly with Chapter 5, and parts of Chapters 2, 3, and 6. One primarily interested in the electromagnetics of the lower frequencies will, on the other hand, find the first six chapters of most value. These chapters may, in fact, be considered as a fairly complete discussion of the electromagnetics underlying electrical engineering up to the higher frequencies, and including an introduction to them.

The system of units used throughout is the mks system of practical units, which has fortunately received common acceptance during the past few years for engineering presentations of electromagnetic subjects. However, the laws are first introduced in the older systems of units (electrostatic and electromagnetic) so that students may have enough

familiarity with these to use effectively the many valuable books and articles employing the older systems.

The authors wish to express their thanks for suggestions, corrections, and other valuable help in the preparation of this textbook to many students and members of the staff of the Advanced Engineering Program of the General Electric Company, and especially to Mr. J. F. McAllister.

SIMON RAMO
JOHN R. WHINNERY

Schenectady, New York
March, 1944

CONTENTS

1

OSCILLATION AND WAVE FUNDAMENTALS

1.01 Introduction

This text is concerned with electromagnetics, particularly that underlying oscillations and waves. Before introducing the laws of electricity and magnetism for serious study, it will be necessary to discuss some ideas and mathematics that have to do with oscillations and waves generally. This will be done by using simple circuits and conventional uniform transmission lines as examples. When this is done, the objective is not to present the theory of circuits and lines as such. Indeed the theory underlying both comprises a good part of the text. The purpose of this chapter is to illustrate (and for some readers to review) a point of view toward oscillations and waves needed for the rest of the text. Specifically the objectives are:

1. To present a clear picture of the energy relations in oscillating systems.
2. To point out criteria relating energy properties of a system to band width, impedance, etc., for later comparison purposes with cavity resonators.
3. To clarify the concepts of waves, particularly in regard to such properties as phase velocity, reflection, and characteristic impedance.
4. To point out common properties of transmission lines according to the conventional distributed constant approach for later comparison with properties of waves in space and in wave guides.
5. To present or review some fundamental mathematics necessary for the study of oscillations and waves throughout the book.
6. To develop approximate methods of analysis based upon the physical picture of the phenomena, so that these may be used in the later, more difficult problems.

SIMPLE CIRCUITS AS EXAMPLES OF OSCILLATING SYSTEMS

1.02 Free Oscillations in an Ideal Simple Circuit

Let us start with the simplest possible circuit for electrical oscillations, an ideal condenser connected across an ideal inductance. Consider first free oscillations, assuming that an amount of energy was supplied to

the combination at some instant (for example, by placing a charge on the condenser) and that from that time on there is no connection to the outside. Energy may be stored in the system in two forms:

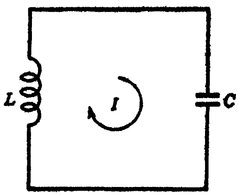


FIG 1 02.

1. Magnetic energy in the inductance. This is analogous to kinetic energy in mechanics and has the value

$$U_L = \tfrac{1}{2}LI^2 \qquad [1]$$

where I is the current flowing through the inductance L.

2. Electric energy in the capacitance. This is analogous to potential energy in mechanics and has the value

$$U_C = \tfrac{1}{2}CV^2 \qquad [2]$$

where V is the voltage across the condenser C.

The presence of energy in the condenser implies a voltage across the condenser, and a consequent rate of change of current and stored magnetic energy in the inductance. Similarly, the presence of magnetic energy requires a current flowing in the inductance, and a consequent rate of change of voltage and stored electric energy in the condenser. We are led then to expect oscillations, since the presence of energy in one form requires a rate of change of energy in the other. It is also necessary that the total energy in the system must be a constant, the same at all instants, since there is no connection to the outside and ideal dissipationless conditions are assumed.

Before going further with purely physical reasoning, let us write an equation for the instantaneous current in the circuit. By Kirchhoff's laws, the sum of the induction voltage, $L\dfrac{dI}{dt}$, and the condenser voltage $\dfrac{q}{C}$ must be zero.

$$L\frac{dI}{dt} + \frac{1}{C}\int I\,dt = 0 \qquad [3]$$

If this equation is differentiated with respect to time, it becomes a true differential equation.

$$L\frac{d^2I}{dt^2} + \frac{I}{C} = 0$$

or

$$\frac{d^2I}{dt^2} = -\frac{I}{LC} \qquad [4]$$

The differential Eq. (4) is called the simple harmonic motion equation. This is probably the simplest and most common of all differential equations. It will probably be so familiar that the reader will wonder why we do not immediately write down the answer to the equation. The objectives here, however, are not to obtain answers to these simple and well-known problems, but rather to freshen up old techniques and to develop new ones for the much more interesting problems that lie ahead.

1.03 Solution to the Simple Harmonic Motion Equation by Assumed Series

The differential equation to be solved is 1.02(4);

$$\frac{d^2I}{dt^2} = -\frac{I}{LC} \tag{1}$$

The method to be shown first for solution of this simple differential equation is one which will be necessary for later less familiar equations, such as the Bessel equation. The method merely recognizes that the solution to a given differential equation can often be expanded in a power series. Conversely, we may assume a general power series at the beginning, and determine what form its coefficients must have if the series is to be a solution for the equation. The required form may be recognizable as the expansion for a known function. At any rate the entire series, if convergent, may always be used as the solution.

Let us then assume that the solution to (1) will be some series of the form

$$I = a_0 + a_1t + a_2t^2 + a_3t^3 + a_4t^4 + \cdots \tag{2}$$

Differentiating,

$$\frac{dI}{dt} = a_1 + 2a_2t + 3a_3t^2 + 4a_4t^3$$

And again,

$$\frac{d^2I}{dt^2} = 2 \cdot 1a_2 + 3 \cdot 2a_3t + 4 \cdot 3a_4t^2 + \cdots$$

These series forms may be substituted in (1) to determine the requirements on the coefficients in order that the series may satisfy that equation.

$$2 \cdot 1a_2 + 3 \cdot 2a_3t + 4 \cdot 3a_4t^2 + 5 \cdot 4a_5t^3 + 6 \cdot 5a_6t^4 + \cdots$$
$$= -\frac{1}{LC}(a_0 + a_1t + a_2t^2 + a_3t^3 + a_4t^4 + \cdots)$$

It may not be obvious at once, but a little study shows that if the above equation is to be true for all values of t, coefficients of like powers of t must be equal on the two sides of the equation. That is,

$$a_2 = -\frac{a_0}{2 \cdot 1LC}$$

$$a_3 = -\frac{a_1}{3 \cdot 2LC}$$

$$a_4 = -\frac{a_2}{4 \cdot 3LC} = \frac{a_0}{4!\,(LC)^2}$$

$$a_5 = -\frac{a_3}{5 \cdot 4LC} = \frac{a_1}{5!\,(LC)^2}$$

$$a_6 = -\frac{a_4}{6 \cdot 5LC} = -\frac{a_0}{6!\,(LC)^3}$$

$$a_7 = -\frac{a_5}{7 \cdot 6LC} = -\frac{a_1}{7!\,(LC)^3}$$

and generalizing,

$$a_{2n} = -\frac{a_{2n-2}}{(2n)(2n-1)LC} = \frac{(-1)^n a_0}{(2n)!\,(LC)^n}$$

$$a_{2n+1} = -\frac{a_{2n-1}}{(2n+1)(2n)LC} = \frac{(-1)^n a_1}{(2n+1)!\,(LC)^n}$$

Notice that the requirements placed upon the constants of the series by substituting in the differential equation have related all constants either to a_0 or a_1, but there is nothing relating these two to each other or to anything else. This seems promising, for two independent solutions and two arbitrary constants are required for a second degree differential equation. Let us write now the assumed series (2), using these constants.

$$I = a_0\left[1 - \frac{t^2}{2!\,LC} + \frac{t^4}{4!\,(LC)^2} - \frac{t^6}{6!\,(LC)^3} + \cdots\right] \tag{3}$$
$$+ (a_1\sqrt{LC})\left[\frac{t}{(LC)^{1/2}} - \frac{t^3}{3!\,(LC)^{3/2}} + \frac{t^5}{5!\,(LC)^{5/2}} - \frac{t^7}{7!\,(LC)^{7/2}} + \cdots\right]$$

Comparison with any tables of series shows that the first quantity in brackets has the form of the series expansion for a cosine function and

the second for a sine. That is,

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots \qquad [4]$$

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots \qquad [5]$$

So (3) may be written

$$I = a_0 \cos\left(\frac{t}{\sqrt{LC}}\right) + a_1\sqrt{LC}\sin\left(\frac{t}{\sqrt{LC}}\right)$$

Since a_1 is arbitrary, the entire quantity $a_1\sqrt{LC}$ may be replaced by C_2 to stress the point that it is an arbitrary constant. Let us at the same time replace a_0 by C_1 and define

$$\omega_0 = \frac{1}{\sqrt{LC}} \qquad [6]$$

Then

$$I = C_1 \cos \omega_0 t + C_2 \sin \omega_0 t \qquad [7]$$

This expression is a solution to the differential equation. It has two independent functions and two arbitrary constants. All is now known except the values of these constants. These cannot be determined until more information is given about the manner of starting oscillations in the circuit.

1.04 Solution of the Simple Harmonic Motion Equation by Assumed Sinusoids

The simple harmonic motion differential equation has been solved by assuming a series solution, determining the form required of that series by the differential equation, and identifying the resulting series as a sinusoidal function. Now, we might have guessed at the beginning that the solution would have been of a sinusoidal form. Although the frequency was not known, we might have assumed a solution of the form

$$I = C_1 \cos \omega_0 t + C_2 \sin \omega_0 t \qquad [1]$$

where ω_0 has to be determined. It remains to be seen if a function of this form can satisfy the differential equation which is

$$\frac{d^2I}{dt^2} = -\frac{I}{LC} \qquad [2]$$

If I is given by (1)

$$\frac{dI}{dt} = -\omega_0(C_1 \sin \omega_0 t - C_2 \cos \omega_0 t)$$

$$\frac{d^2I}{dt^2} = -\omega_0^2(C_1 \cos \omega_0 t + C_2 \sin \omega_0 t) \qquad [3]$$

Substitute (1) and (3) in (2)

$$-\omega_0^2(C_1 \cos \omega_0 t + C_2 \sin \omega_0 t) = -\frac{1}{LC}(C_1 \cos \omega_0 t + C_2 \sin \omega_0 t)$$

If

$$\omega_0^2 = \frac{1}{LC} \qquad [4]$$

the equation is satisfied. This value of ω_0^2 is exactly that defined in Eq. 1.03(6).

Thus it is demonstrated that if we can guess the form of a solution to a differential equation, substitution of this form into the equation will determine whether or not it is a solution and will give values for any non-arbitrary constants, such as ω_0 above. This method is one of the most useful for solution of differential equations in engineering.

1.05 Solution of the Simple Harmonic Motion Equation by Assumed Exponentials

As a final attack on the differential equation for simple harmonic motion we shall attempt a solution in terms of exponentials. The wisdom of this will shortly be demonstrated. Suppose we try

$$I = A_1e^{pt} + A_2e^{-pt} \qquad [1]$$

then

$$\frac{d^2I}{dt^2} = p^2(A_1e^{pt} + A_2e^{-pt})$$

Substitute these in Eq. 1.03(1)

$$p^2(A_1e^{pt} + A_2e^{-pt}) = -\frac{1}{LC}(A_1e^{pt} + A_2e^{-pt})$$

or

$$p^2 = -\frac{1}{LC}$$

$$p = j\sqrt{\frac{1}{LC}} = j\omega_0$$

where $j = \sqrt{-1}$.

This substitution indicates that (1) is a solution to the simple harmonic motion equation, provided that $p = j\omega_0$,

$$I = A_1 e^{j\omega_0 t} + A_2 e^{-j\omega_0 t} \tag{2}$$

Next let us remind ourselves of the identities

$$e^{jx} = \cos x + j \sin x \tag{3}$$

$$e^{-jx} = \cos x - j \sin x \tag{4}$$

These are most conveniently verified by considering the series expansion for an exponential.

$$e^y = 1 + y + \frac{y^2}{2!} + \frac{y^3}{3!} + \cdots$$

so

$$e^{jx} = 1 + jx - \frac{x^2}{2!} - j\frac{x^3}{3!} + \cdots$$

$$= \left(1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots\right) + j\left(x - \frac{x^3}{3!} + \frac{x^5}{5!} + \cdots\right)$$

By comparing with Eq. 1.03(4) and Eq. 1.03(5) these latter series are quickly identified as those for cosine and sine respectively, thus verifying (3). The corresponding demonstration for (4) is identical to this.

If the identities (3) and (4) are substituted in (2),

$$I = (A_1 + A_2) \cos \omega_0 t + j(A_1 - A_2) \sin \omega_0 t$$

Since A_1 and A_2 are both arbitrary, this may be written exactly in the previous forms,

$$I = C_1 \cos \omega_0 t + C_2 \sin \omega_0 t \tag{5}$$

For many purposes it will be convenient to use the solution in the form of (2) instead of changing to (5). This use of exponentials to replace sinusoids will be the subject of later discussion.

Problem 1.05. Show that an alternative expression equivalent to Eq. 1.05(2) or Eq. 1.05(5) is

$$I = A \cos (\omega_0 t + \phi)$$

Relate A and ϕ to C_1 and C_2.

1.06 Natural Oscillations with Losses — Approximate Method

The circuit analyzed previously was ideal. Suppose we now wish to consider the effect of the finite losses which must of necessity be present in the circuit. As will be shown in the next section, it is a simple matter

to include these in the circuit equations rigorously, yet let us first use physical knowledge to develop an approximate method which will give the first order effect of the losses, provided that losses are small. The point of view will be extremely useful in later analyses of cavity resonators and wave guides.

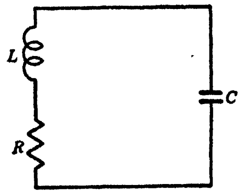


FIG. 1.06.

If losses are small, physical intuition tells us that the natural period of oscillation will be changed little, and over a short period of time the solution will be very nearly that for the ideal circuit. The major correction will be a long-time decrease in the amplitude of oscillation due to the energy lost.

It is common experience to find exponential changes for a physical quantity which decreases (or increases) at a rate proportional to the amount of that quantity present. The power loss, or rate of energy decrease, for this example, is proportional to the amount of energy in the system. It would consequently be reasonable to expect an exponential damping factor to appear in the expressions for currents and voltages. As a first order correction, the expression for current obtained previously (Prob. 1.05) might be assumed to be multiplied by some negative exponential

$$I = Ae^{-\alpha t} \cos (\omega_0 t + \phi) \qquad [1]$$

The energy in the circuit may be calculated at an instant when it is all in the inductance

$$U = \tfrac{1}{2}L(I_{\max})^2 = \frac{LA^2}{2} e^{-2\alpha t} \qquad [2]$$

Within the limits of the assumption of relatively small losses, the negative rate of change of this stored energy over several cycles is merely the average power loss.

$$-\frac{dU}{dt} = W_L \qquad [3]$$

From (2),

$$\frac{dU}{dt} = -2\alpha \frac{LA^2}{2} e^{-2\alpha t} = -2\alpha U \qquad [4]$$

So, by combining (3) and (4),

$$\alpha = \frac{W_L}{2U} \qquad [5]$$

Define the quality factor or Q of the circuit as the quantity

$$Q = \frac{\omega_0 \text{ (Energy stored in circuit)}}{\text{Average power loss}} = \frac{\omega_0 U}{W_L} \tag{6}$$

$$= \frac{\pi\text{(Energy stored in circuit)}}{\text{Energy lost per half cycle}} \tag{7}$$

Then (5) may be written

$$\alpha = \frac{\omega_0}{2Q} \tag{8}$$

The exponential decay is thus expressible in terms of the quantity Q. The damping is also described sometimes as a logarithmic decrement, which is the relative amount by which the amplitude of oscillation decreases in one period.

$$\delta = \frac{Ae^{-\alpha t} - Ae^{-\alpha(t+T)}}{Ae^{-\alpha t}} = 1 - e^{-\alpha T} \cong \alpha T$$

provided αT is small compared with unity, or

$$\delta = \frac{\omega_0}{2Q} T = \frac{2\pi f_0}{2Q} \times \frac{1}{f_0} = \frac{\pi}{Q} \tag{9}$$

Finally, let us interpret these results for a circuit with losses distributed as in Fig. 1.06. The current flow through the series combination of R and L is expressed by

$$I = A \cos(\omega_0 t + \phi)$$

(neglecting any exponential damping for a few cycles). The energy stored in the circuit is the maximum energy in the inductance,

$$U = \frac{L}{2} A^2$$

and the average power loss in resistance R is

$$W_R = \tfrac{1}{2} R (I_{\max})^2 = \frac{RA^2}{2}$$

So Q, defined by (6), is

$$Q_L = \frac{\omega_0(LA^2)}{RA^2} = \frac{\omega_0 L}{R} \tag{10}$$

This is the familiar expression for Q used to describe the excellence of an inductance, $\omega L/R$ calculated at resonance. It is to be used in (8) or (9) to give attenuation constant or logarithmic decrement.

Problem 1.06. (*a*) If losses are present owing to a conductance $G = 1/R_1$ shunted across the condenser instead of a series resistance in the inductance, show that the Q to use in the general expressions Eq. 1.06(8) and Eq. 1.06(9) is

$$Q_C = \frac{\omega_0 C}{G} = \frac{R_1}{\omega_0 L}$$

(*b*) If losses arise from both series resistance in L and shunt conductance across C, demonstrate that the Q to use in the general expressions may be found from the individual Q's defined previously.

$$\frac{1}{Q} = \frac{1}{Q_L} + \frac{1}{Q_C}$$

1.07 Natural Oscillations with Losses; Solution from Circuit Equations

The exact solution to the circuit of Fig. 1.06 will now be obtained to check the approximate results of the previous article.

$$L\frac{dI}{dt} + RI + \frac{1}{C}\int I\,dt = 0$$

is the exact equation of the circuit. Differentiating,

$$L\frac{d^2I}{dt^2} + R\frac{dI}{dt} + \frac{I}{C} = 0 \qquad [1]$$

Following the method of Art. 1.05, assume a solution of exponential form,

$$I = Ae^{pt} \qquad [2]$$

If this is substituted in (1) and the resulting equation is solved for p, it is found that

$$p = -\frac{R}{2L} \pm \sqrt{\left(\frac{R}{2L}\right)^2 - \frac{1}{LC}} \qquad [3]$$

Since for low-loss circuits $(R/2L)^2$ will be less than $1/LC$, it will be convenient to write (3) as

$$p = -\frac{R}{2L} \pm \frac{j}{\sqrt{LC}}\sqrt{1 - \frac{R^2C}{4L}} = -\alpha \pm j\omega_0' \qquad [4]$$

where

$$\alpha = \frac{R}{2L} = \frac{\omega_0}{2Q} \qquad [5]$$

$$\omega_0' = \frac{1}{\sqrt{LC}}\sqrt{1 - \frac{R^2C}{4L}} = \omega_0\sqrt{1 - \left(\frac{1}{2Q}\right)^2} \qquad [6]$$

Q denotes $\omega_0 L/R$ as in Eq. 1.06(10), and ω_0 is $1/\sqrt{LC}$.

The two possible values of p from (4) supply the two independent solutions needed for the second degree differential equation. Substitute these in (2).

$$I = A_1 e^{(-\alpha + j\omega_0')t} + A_2 e^{(-\alpha - j\omega_0')t}$$
$$= e^{-\alpha t}[A_1 e^{j\omega_0' t} + A_2 e^{-j\omega_0' t}]$$

By substitutions similar to those of Art. 1.02, an alternative expression is

$$I = e^{-\alpha t}[C_1 \cos \omega_0' t + C_2 \sin \omega_0' t] \qquad [7]$$

A comparison with the approximate analysis of Art. 1.06 shows that the same damping coefficient (5) is obtained. The natural frequency is different from ω_0 by (6), but this difference is small for low-loss (high-Q) circuits.

Problem 1.07. Obtain exact results for the cases solved approximately in Probs. 1.06(a), (b), showing for these also that Q may be used as an indication of the usefulness of the approximate results.

1.08 Forced Oscillations in an Ideal L-C Circuit

In previous examples, it was assumed that oscillations in the simple resonant circuit were free oscillations caused only by an initial deposit of energy in the circuit. In most practical cases, however, the circuit is continuously excited by a source of sinusoidal voltage. As the first example of such forced oscillations, consider the loss-free parallel L-C circuit excited by a sinusoidal voltage of constant magnitude (Fig. 1.08). The total current flow from the source is the sum of currents in the two impedances. The equations for these two currents are

$$L \frac{dI_1}{dt} = V \sin \omega t \qquad [1]$$

$$\frac{1}{C}\int I_2\, dt = V \sin \omega t \qquad [2]$$

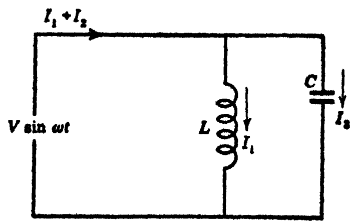


FIG. 1.08.

Current may be obtained from (1) by integrating directly and from (2) by differentiating.

$$LI_1 = -\frac{V}{\omega} \cos \omega t + C_1 \qquad [3]$$

$$\frac{I_2}{C} = \omega V \cos \omega t \qquad [4]$$

The constant term in (3) merely represents a possible constant D-C

term flowing through the inductance, which is of no interest to the A-C problem so long as constant elements (linear systems) are assumed. Thus the total current

$$I = I_1 + I_2 = V\left(\omega C - \frac{1}{\omega L}\right)\cos \omega t \tag{5}$$

The above relations, of course, check the well-known behavior of simple circuits. The current in the inductance has a phase lag of 90° with respect to its voltage, whereas the current in the capacitance has a 90° phase lead with respect to the voltage. The total current is leading (the total circuit acts as a capacitance) if $\omega C > (1/\omega L)$ and is lagging (total circuit acts as an inductance) if $(1/\omega L) > \omega C$. If $\omega C = 1/\omega L$ there is no current to be supplied by the source; under this condition the current flow to the inductance is at every instant exactly equal and opposite to the current flow to the capacitance. The frequency for which this condition occurs is the natural frequency found previously,

$$\omega C = \frac{1}{\omega L} \quad \text{or} \quad \omega = \frac{1}{\sqrt{LC}} = \omega_0 \tag{6}$$

At this natural frequency the energy inside the system is a constant and merely passes back and forth from inductance to capacitance, and no energy need be supplied by the source at any instant of time. For a frequency lower than this resonance frequency, the maximum energy stored in the inductance is greater than the maximum energy stored in the capacitance, so that this excess energy must be supplied from the source during one part of the cycle, but will be delivered back to it unharmed during another part. This excess reactive energy from the inductance makes the circuit appear as an inductive load to the source. Similarly, for frequencies greater than the resonant frequency, the maximum energy in the capacitance is greater than the maximum energy in the inductance, and the excess reactive energy that must be supplied to the capacitance causes the circuit to appear as a capacitive load to the source.

At the resonant frequency, the energy stored in the circuit is the maximum energy of the capacitance, or the maximum energy stored in the inductance, since both are equal.

$$U = \tfrac{1}{2}CV^2$$

For later use, let us write this from (6) in terms of ω_0.

$$U = \frac{V^2}{2\omega_0\sqrt{\frac{L}{C}}} \tag{7}$$

1.09 Approximate Input Impedance at Resonance with Losses Considered

If the parallel circuit has losses in the coil or condenser, these may be taken into account from physical consideration of the energy relations, before attempting an exact analysis by the circuit equations.

At resonance, the energy stored in the tuned circuit is given by Eq. 1.08(7). From the definition of Q given in Eq. 1.06(6), the power loss at resonance is

$$W_L = \frac{\omega_0 U}{Q} = \frac{V^2}{2Q}\sqrt{\frac{C}{L}} \qquad [1]$$

The source must supply to the circuit this amount of power. The circuit then looks like a high-resistance R_i of value such that

$$W = \frac{V^2}{2R_i} \qquad [2]$$

By comparing with (1),

$$R_i = Q\sqrt{\frac{L}{C}} \qquad [3]$$

The approximations of reasonably low losses will be recognized in the above reasoning, for we have taken the expression for energy stored as that developed from the loss-free case. In this picture, the major part of the energy is stored in the circuit and passes back and forth from the inductance to the capacitance. Only the small amount of power lost in the process need be supplied by the source. The resulting current flow to supply this loss component causes the circuit to have a high but finite input impedance in place of the infinite input impedance found previously.

Problem 1.09. (*a*) Write alternative forms for Eq. 1.09(3) in terms of circuit reactances at resonance.

(*b*) Write Eq. 1.09(3) in terms of: a series resistance in L, a shunt resistance across C, both series and shunt losses.

1.10 Approximate Input Impedance near Resonance

The physical reasoning may be extended to give the behavior of the circuit approximately for a small departure from resonance. First it may be concluded that the major change will appear as a reactive component added to the admittance as frequency is changed to a value such that the capacitive and inductive reactive currents no longer cancel. To a first approximation, the input power supplied will be constant, so that the conductive portion of the admittance may be considered con-

stant and equal to that calculated at resonance in Art. 1.09. We justify this by recognizing that any loss entering from a parallel conductance will not change at all with frequency, and although that arising from a resistance in series with inductance will change with frequency, this is a uniform change, not comparable with the change in the differences of large quantities which affects the reactive current. Then

$$G_i = \frac{1}{R_i} = \frac{1}{Q}\sqrt{\frac{C}{L}} \qquad [1]$$

The susceptance portion of the admittance is approximately that calculated without losses.

$$S_i = \left(\omega C - \frac{1}{\omega L}\right) = \omega C\left(1 - \frac{\omega_0^2}{\omega^2}\right)$$

Let $\omega = \omega_0(1 + \delta)$. Then

$$S_i = \sqrt{\frac{C}{L}}\left[(1 + \delta) - \frac{1}{1 + \delta}\right]$$

For frequencies near resonance, $\delta \ll 1$,

$$(1 + \delta)^{-1} \cong 1 - \delta$$

and

$$S_i \cong 2\delta\sqrt{\frac{C}{L}} \qquad [2]$$

By comparing (1) and (2), it is evident that the frequency shift for which the susceptance becomes equal to the conductance, a common measure of circuit "sharpness," is

$$\delta = \frac{1}{2Q} \qquad [3]$$

The Q of the circuit is consequently identified with the band width or sharpness of the circuit. At a frequency such that susceptance and conductance are equal, $G_i = S_i$, the magnitude of input admittance

$$|Y| = \sqrt{G_i^2 + S_i^2} = \frac{\sqrt{2}}{Q}\sqrt{\frac{C}{L}} \qquad [4]$$

In terms of impedance, the impedance at this frequency is $1/\sqrt{2}$ its value at resonance.

USE OF COMPLEX EXPONENTIALS

1.11 Solution of the Circuit Differential Equation in Terms of Complex Exponentials

The approximate results of the previous articles for the circuit relations when dissipation is included will now be verified by direct solution of the differential equation of the circuit. If a voltage $V \cos \omega t$ is applied to a circuit containing R, L, and C in series, the equation to be solved is

$$L\frac{dI}{dt} + RI + \frac{1}{C}\int I\,dt = V \cos \omega t \qquad [1]$$

But (see Eqs. 1.05(3), (4))

$$\cos \omega t = \frac{e^{j\omega t} + e^{-j\omega t}}{2} \qquad [2]$$

If we assume the current to have the steady state solution

$$I = Ae^{j\omega t} + Be^{-j\omega t} \qquad [3]$$

The result of substituting in (1) is

$$j\omega L(Ae^{j\omega t} - Be^{-j\omega t}) + R(Ae^{j\omega t} + Be^{-j\omega t}) + \frac{1}{j\omega C}(Ae^{j\omega t} - Be^{-j\omega t}) = \frac{V}{2}[e^{j\omega t} + e^{-j\omega t}] \qquad [4]$$

Following previous reasoning, this equation can be true for all values of time only if coefficients of $e^{j\omega t}$ are the same on both sides of the equation, and similarly for $e^{-j\omega t}$.

$$A\left[R + j\left(\omega L - \frac{1}{\omega C}\right)\right] = \frac{V}{2} \qquad [5a]$$

$$B\left[R - j\left(\omega L - \frac{1}{\omega C}\right)\right] = \frac{V}{2} \qquad [5b]$$

The complex quantity in the bracket of (5a) may be called Z and written in its equivalent form

$$Z = R + j\left(\omega L - \frac{1}{\omega C}\right) = |Z|e^{j\psi}$$

where

$$|Z| = \sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2} \qquad [6]$$

and

$$\psi = \tan^{-1} \frac{\left(\omega L - \dfrac{1}{\omega C}\right)}{R} \qquad [7]$$

Similarly,

$$R - j\left(\omega L - \frac{1}{\omega C}\right) = |Z|e^{-j\psi}$$

Then,

$$A = \frac{V}{2|Z|} e^{-j\psi}$$

$$B = \frac{V}{2|Z|} e^{j\psi}$$

(A and B are conjugates: they have the same real parts and equal and opposite imaginary parts.) Substituting in (3),

$$I = \frac{V}{|Z|}\left[\frac{e^{j(\omega t-\psi)} + e^{-j(\omega t-\psi)}}{2}\right] \qquad [8]$$

By comparing with (2),

$$I = \frac{V}{|Z|} \cos (\omega t - \psi) \qquad [9]$$

This final result gives the desired magnitude and phase angle of the current with respect to the applied voltage. That information is contained in either of the constants A or B, and no information is given in one which is not in the other. B is of necessity the conjugate of A, since this is the only way in which the two may add up to a real current, and the final exact answer for current must be real. It follows that half of the work was unnecessary. We could have started only with $Ve^{j\omega t}$ in place of the two-term expression which is exactly equivalent to $V \cos \omega t$. For current, there would then be only

$$I = \frac{V}{|Z|} e^{j(\omega t-\psi)} \qquad [10]$$

Although this cannot actually be the expression for current, since it is a complex and not a real quantity, it contains all the information we wish to know: magnitude of current, $V/|Z|$, and its phase with respect to applied voltage, ψ. The procedure for obtaining the steady state solution to differential equations with applied sinusoids may then be summarized as below. The steps will be verified by a check of the previous steps in the exact solution.

1. Write sinusoidal applied voltage as $Ve^{j\omega t}$
2. Replace dI/dt by $j\omega I$, $\int I\,dt$ by $I/j\omega$.
3. Solve for current in the resulting algebraic equation. The answer will in general be complex.
4. By writing the complex current in its form showing magnitude and phase angle, the desired information is obtained.

The method, as outlined, is used by electrical engineers so generally that they have come to think quite naturally in terms of the complex impedance, Z, of the circuit, often without recalling that they are in reality solving a differential equation for its steady state solution.

Problem 1.11. Show that Eq. 1.11 (10) need not be regarded simply as a representation of the true expression for current but may actually be an exact expression, provided only that it be understood that I is the real part of the complex expression $\dfrac{V}{|Z|}e^{j(\omega t-\psi)}$. Sometimes this relationship is expressed by writing

$$I = Re\left\{\frac{V}{|Z|}e^{j(\omega t-\psi)}\right\}$$

1.12 Use of Complex Quantities in Power Calculations

The preceding article demonstrated the basis for using $e^{j\omega t}$ as a representation for $\cos \omega t$. The consequent simplification of problems involving impressed sinusoidal quantities will be apparent throughout the book. However, we must remember one trap that awaits us if we use this notation improperly in non-linear expressions. For linear equations, one may use $e^{j\omega t}$ in place of the equivalent cosine or sine term in a completely straightforward fashion, interpreting information of magnitudes and phase angles as demonstrated in the previous article. More care must be exercised for non-linear expressions, the most common of which arises in the calculation of instantaneous power, requiring a product of terms.

Given a sinusoidal voltage across an impedance

$$V = V_m \cos(\omega t - \phi_1) \quad [1]$$

and a sinusoidal current flow through the impedance

$$I = I_m \cos(\omega t - \phi_2) \quad [2]$$

These are properly represented in complex notation as follows.

$$V = V_m e^{j(\omega t-\phi_1)} \quad [3]$$

$$I = I_m e^{j(\omega t-\phi_2)} \quad [4]$$

The power flow at any instant is given by

$$W_i = VI$$

There is certainly a strong temptation to multiply the expressions (3) and (4) together to give

$$W = V_m I_m e^{j(\omega t - \phi_1)} e^{j(\omega t - \phi_2)} = V_m I_m e^{j(2\omega t - \phi_1 - \phi_2)} \qquad [5]$$

This is the complex representation for a quantity

$$W = V_m I_m \cos (2\omega t - \phi_1 - \phi_2) \qquad [6]$$

This result is incorrect. It must be incorrect since the average power in (6) is zero, regardless of the phase angle ϕ_1 or ϕ_2. The correct expression for instantaneous power is certainly given by multiplying (1) and (2)

$$W_i = VI = V_m I_m \cos (\omega t - \phi_1) \cos (\omega t - \phi_2)$$

but

$$\cos A \cos B = \tfrac{1}{2}[\cos (A - B) + \cos (A + B)]$$

so

$$W_i = \frac{V_m I_m}{2} \cos (\phi_1 - \phi_2) + \frac{V_m I_m}{2} \cos (2\omega t - \phi_1 - \phi_2) \qquad [7]$$

Equation (7) differs mainly from the result of (6) in that it does have the average power in its familiar form: product of voltage, current, and cosine of the phase angle. ($\frac{1}{2}$ appears since V_m and I_m are peak values, not rms.)

The use of (3) and (4) in the product expression for power is incorrect simply because the expressions (3) and (4) are not the true expressions for voltage and current, but merely representations for them. We would not have invited such a difficulty had the true mathematical equivalents (Prob. 1.11) been written:

$$V = Re\,[V_m e^{j\omega t}]$$

$$I = Re\,[I_m e^{j(\omega t - \phi)}]$$

Then instantaneous power,

$$W_i = \{Re\,[V_m e^{j\omega t}]\}\{Re\,[I_m e^{j(\omega t - \varphi)}]\}$$

To evaluate W_i, we could write these as cosines and proceed as before. The foregoing may seem to be an argument for retaining the notation $Re\,[e^{j\omega t}]$ but this will not be done because of the obvious unwieldiness of the expressions. We shall use only $e^{j\omega t}$ with the real operator under-

stood, or merely say that this is a representation for $\cos \omega t$. Care must then be used for any products of these exponentials.

The exact expression for instantaneous power may be written in complex notation. For the following demonstration, let us denote complex quantities by a wavy line above the symbol, and conjugates by an asterisk.

If
$$\tilde{V} = V_m e^{j\phi_1}$$
and
$$\tilde{V}^* = V_m e^{-j\phi_1}$$
then
$$\tilde{V}\tilde{I}^* = V_m I_m e^{j(\phi_1 - \phi_2)}$$
and
$$\tilde{V}^*\tilde{I} = V_m I_m e^{-j(\phi_1 - \phi_2)}$$
Then the exact equivalent of (7) in complex notation is
$$W = \tfrac{1}{2} Re \left\{ \tilde{V}\tilde{I}^* + [\tilde{V}e^{j\omega t}][\tilde{I}e^{j\omega t}] \right\} \qquad [8]$$
and the average power, or constant component of this,
$$W_{av} = \tfrac{1}{2} Re\,[\tilde{V}\tilde{I}^*] = \tfrac{1}{2} Re\,[\tilde{V}^*\tilde{I}] \qquad [9]$$

FOURIER SERIES

1.13 Circuits with Non-Sinusoidal Periodic Voltages

All forced oscillations studied so far have consisted of sinusoids. Consider a more general oscillation which is periodic, returning once each cycle to any selected reference, or stated mathematically,
$$f(t) = f(t - T)$$
This might be of any arbitrary form, such as is indicated by Fig. 1.13. Such a wave shape of voltage, if applied to a circuit, will act to that circuit as a superposition of a group of pure sinusoidal voltages. The wave may be replaced by a fundamental and its harmonics. The method

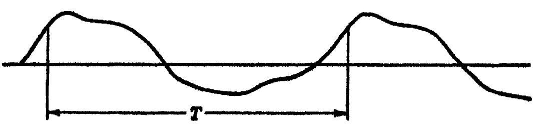


FIG. 1.13. Periodic wave of arbitrary shape.

of finding the amplitude of these is the very neat method of Fourier analysis, and the theorem that proves the truth of the foregoing statements is the Fourier theorem, which it is assumed the reader has agreed

with in another study. What follows here is not a proof of the validity of a Fourier series expansion for a general periodic function, but merely a demonstration which shows the manner of obtaining the coefficients. This will be extremely useful when we later add up series to represent known functions along boundaries in field problems.

By admitting the possibility of writing the periodic function $f(t)$ as a series of sinusoids consisting of a fundamental and its harmonics,

$$f(t) = a_0 + a_1 \cos \omega t + a_2 \cos 2\omega t + a_3 \cos 3\omega t + \cdots$$
$$+ b_1 \sin \omega t + b_2 \sin 2\omega t + b_3 \sin 3\omega t \cdots \quad [1]$$

At the moment, the coefficients have not been determined. The manner of finding them is based upon the so-called orthogonality property of sinusoids. This property indicates that the integral of the product of any two sinusoids of different frequencies, over an interval in which they are commensurate (for example, from $-\pi$ to π, or 0 to 2π) shall be zero. That is,

$$\int_{0\ m\neq n}^{2\pi} \cos mx \cos nx\, dx = 0$$
$$\int_{0\ m\neq n}^{2\pi} \sin mx \sin nx\, dx = 0 \quad [2]$$
$$\int_{0\ m\neq n \text{ or } m=n}^{2\pi} \sin mx \cos nx\, dx = 0$$

However,

$$\int_0^{2\pi} \cos^2 mx\, dx = \int_0^{2\pi} \sin^2 mx\, dx = \pi \quad [3]$$

Thus if each term in (1) is multiplied by $\cos n\omega t$, and integrated from 0 to 2π, every term on the right will be zero except that term containing a_n. That is

$$\int_0^{2\pi} f(t) \cos n\omega t\, d(\omega t) = \int_0^{2\pi} a_n \cos^2 n\omega t\, d(\omega t)$$

By (3), the integral on the right has the value $a_n\pi$, or

$$a_n = \frac{1}{\pi} \int_0^{2\pi} f(t) \cos n(\omega t)\, d(\omega t) \quad [4]$$

Similarly, to obtain b_n, each term in (1) is multiplied by $\sin n\omega t$ and integrated from 0 to 2π. Then,

$$b_n = \frac{1}{\pi} \int_0^{2\pi} f(t) \sin n(\omega t)\, d(\omega t) \quad [5]$$

Finally, to obtain the constant term a_0, every term is integrated directly over a period, and all terms on the right disappear, except that containing a_0.

$$\int_0^{2\pi} f(t)\, d(\omega t) = \int_0^{2\pi} a_0\, d(\omega t) = 2\pi a_0$$

or

$$a_0 = \frac{1}{2\pi} \int_0^{2\pi} f(t)\, d(\omega t) \qquad [6]$$

This merely states that a_0 is the average of the function $f(t)$.

Before discussing the method in general terms further, let us tie it down by application to a very simple example.

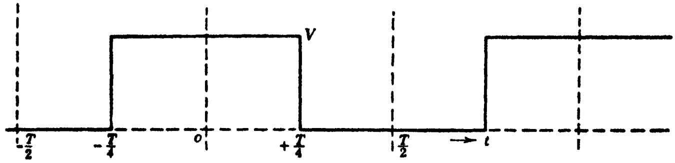


Fig. 1.14. Periodic wave of rectangular shape.

1.14 Fourier Analysis of a Square Wave Voltage

Let us find by Fourier analysis the coefficients of the frequency components in the square wave shape of Fig. 1.14. Voltage is V over half the period T, and zero over the remaining half.

Since

$$\omega = 2\pi f = \frac{2\pi}{T}$$

$$\omega t = \pi$$

when

$$t = \frac{T}{2}$$

The integral 1.11(6) shows that the constant term, a_0, is

$$a_0 = \frac{1}{2\pi} \int_{-\pi}^{+\pi} f(t)\, d(\omega t) = \frac{1}{2\pi} \int_{-\pi/2}^{+\pi/2} V\, d(\omega t) = \frac{V}{2} \qquad [1]$$

This is clearly the average value of the wave. The integral 1.13(5) gives the coefficient b_n.

$$b_n = \frac{1}{\pi} \int_{-\pi}^{+\pi} f(t) \sin n(\omega t)\, d(\omega t) = \frac{1}{\pi} \int_{-\pi/2}^{+\pi/2} V \sin n(\omega t)\, d(\omega t) = 0 \qquad [2]$$

Thus all coefficients of the sine terms are zero. This could have been foreseen by noting that the above function is an even function, that is, $f(-t) = f(t)$, and so could not be made up of any sine terms, which are odd functions.

Finally, the a_n terms, by Eq. 1.13(4),

$$a_n = \frac{1}{\pi}\int_{-\pi}^{+\pi} f(t)\cos n\omega t\, d(\omega t) = \frac{1}{\pi}\int_{-\pi/2}^{+\pi/2} V\cos n\omega t\, d(\omega t)$$

$$a_n = \frac{V}{n\pi}\Big[\sin n(\omega t)\Big]_{-\pi/2}^{+\pi/2} \qquad [3]$$

The value of (3) is zero if n is even, is $+(2V/n\pi)$ if n is 1, 5, 9, etc., and is $-(2V/n\pi)$ if n is 3, 7, 11, etc. Thus the series expansion in sinusoids of the square wave voltage of Fig. 1.14 may be written

$$f(t) = \frac{V}{2} + \frac{2V}{\pi}\left[\cos\omega t - \frac{\cos 3\omega t}{3} + \frac{\cos 5\omega t}{5} - \frac{\cos 7\omega t}{7} + \cdots\right] \qquad [4]$$

The current which flows when such a voltage is applied to a circuit is found by determining the currents due to the individual terms of (4) and superposing these. There will be, in general, a component of current of a frequency corresponding to each frequency component of the Fourier expansion. These, when added, give the wave shape of current. Such a procedure is straightforward and will not be carried further here.

Notice that it requires an infinite number of terms to represent truly the square wave shape of voltage. Often a high degree of approximation to the desired wave shape is obtained when only a finite number of terms is used. However, for functions with sharp discontinuities, many terms may be required near the sharp corners, and the theory of Fourier series shows that the derivative of the series may not even converge to the derivative of the function, although the integral of the series does converge to that of the function.[4]

Problem 1.14. Simplify the general expressions for Fourier coefficients found in Art. 1.13 for:

(a) Even functions of t.
(b) Odd functions of t.
(c) Functions of a variable x, in terms of a period l.

UNIFORM TRANSMISSION LINES AS EXAMPLES OF WAVE SYSTEMS

1.15 The Ideal Transmission Line

To illustrate waves, we shall consider the uniform transmission line. The results developed are of importance themselves, since transmission

lines are used in all modern high-frequency applications. Results will also be used for later comparison with more general electromagnetic wave phenomena. The approach used in this chapter is the conventional one, starting from distributed inductance and capacitance along the line. It is true that this in a sense is jumping ahead of the story, for in a later chapter on guided waves the transmission line differential equations will be derived from rigorous considerations of electromagnetic theory. Nevertheless, the approach to be used here is easy to visualize and is satisfactory for the present purpose.

A transmission line may be made up of parallel wires, of parallel plates, of coaxial conductors, or in general of any two conductors separated by a dielectric material. In conventional analyses, we think in terms of a current flowing in the conductors, equal and opposite in the two conductors if measured at any given transverse plane, and a voltage difference existing between the conductors. The current flow is affected by a distributed series inductance representing the back induced voltage effects of magnetic flux surrounding the conductors; the voltage between conductors acts across a distributed shunt capacitance. There are also loss terms which will be neglected for this first analysis of the ideal case. Incidentally, this does not relegate the results to a position of only academic interest, for many high-frequency transmission line problems have loss terms which are truly negligible.

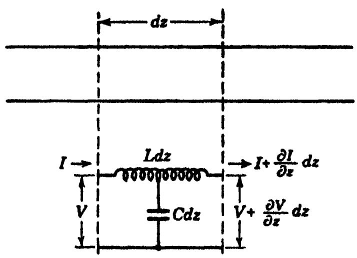


FIG. 1.15.

Consider a differential length of line, dz, including only the distributed inductance, L per unit length, and the distributed capacitance, C per unit length. The length dz then has inductance $L\,dz$ and capacitance $C\,dz$ (Fig. 1.15). The voltage drop or negative change in voltage across this length is then equal to this inductance multiplied by the time rate of change of current. For such a differential length, the voltage change along it at any instant may be written as the length multiplied by the rate of change of voltage with respect to length. Then

$$\text{Voltage change} = \frac{\partial V}{\partial z}\,dz = -(L\,dz)\,\frac{\partial I}{\partial t} \qquad [1]$$

Note that time and space derivatives are written as partial derivatives, since the reference point may be changed in space or time, in completely independent fashion.

Similarly, the decrease in current across the element at any instant is merely that current which is shunted across the distributed capacity. This is given by the capacity multiplied by time rate of voltage. Partial derivatives are again called for.

$$\text{Current change} = \frac{\partial I}{\partial z}\, dz = -(C\, dz)\frac{\partial V}{\partial t} \qquad [2]$$

The length dz may be cancelled in (1) and (2)

$$\frac{\partial V}{\partial z} = -L\frac{\partial I}{\partial t} \qquad [3]$$

$$\frac{\partial I}{\partial z} = -C\frac{\partial V}{\partial t} \qquad [4]$$

Equations (3) and (4) are the fundamental differential equations for the analysis of the ideal transmission line. They may be combined to give equations containing voltage alone or current alone. To accomplish this, differentiate (3) partially with respect to distance, (4) with respect to time.

$$\frac{\partial^2 V}{\partial z^2} = -L\frac{\partial^2 I}{\partial z\, \partial t} \qquad [5]$$

$$\frac{\partial^2 I}{\partial t\, \partial z} = -C\frac{\partial^2 V}{\partial t^2} \qquad [6]$$

Since partial derivatives are the same taken in either order, (6) may be substituted directly in (5).

$$\frac{\partial^2 V}{\partial z^2} = LC\frac{\partial^2 V}{\partial t^2} \qquad [7]$$

This differential equation is known as the wave equation. An exactly similar equation may be obtained in terms of current by differentiating (4) with respect to z, (3) with respect to t, and combining

$$\frac{\partial^2 I}{\partial z^2} = LC\frac{\partial^2 I}{\partial t^2} \qquad [8]$$

1.16 Solutions of the Wave Equation

The differential equation to be solved, Eq. 1.15(7), may be written

$$\frac{\partial^2 V}{\partial z^2} = \frac{1}{v^2}\frac{\partial^2 V}{\partial t^2} \qquad [1]$$

where

$$\frac{1}{v^2} = LC$$

Unlike the differential equations met previously in this chapter, this is a partial differential equation. A direct attack on the equation to yield a general solution is not easy, but a simple check shows that any function whatever in the variable $(t - z/v)$, is a solution. That is,

$$V = F\left(t - \frac{z}{v}\right) \tag{2}$$

is a solution to (1). This may be verified, say, by letting $(t - z/v) = x$. Then

$$\frac{\partial V}{\partial t} = \frac{\partial V}{\partial x}\frac{\partial x}{\partial t} \quad \text{and} \quad \frac{\partial V}{\partial z} = \frac{\partial V}{\partial x}\frac{\partial x}{\partial z}$$

But

$$\frac{\partial x}{\partial t} = 1 \qquad \text{and} \qquad \frac{\partial x}{\partial z} = -\frac{1}{v}$$

so, repeating the process,

$$\frac{\partial^2 V}{\partial t^2} = \frac{\partial^2 V}{\partial x^2} \qquad \text{and} \quad \frac{\partial^2 V}{\partial z^2} = \frac{1}{v^2}\frac{\partial^2 V}{\partial x^2} \tag{3}$$

By comparing the two equations (3), (1) is verified.

It is necessary to show next what is meant by the statement that the solution (2) represents a wave. This may be done in two or three equivalent ways.

1. Note that if the voltage is observed as a function of time in the plane $z = 0$, this is merely $V = F(t + 0) = F(t)$. If we now go to a plane $z = z_1$, we find exactly the same function of time, but delayed by the time z_1/v. That is, the value of V at $z = 0$, and time t will be found at $z = z_1$ at a time $t + (z_1/v)$. The time z_1/v is the time necessary to propagate the effects over the distance z_1 with velocity v.

2. If we were to move along the line with any given point on a traveling wave, the observed value of voltage would be constant. By noting (2), it is seen that this is accomplished if $t - (z/v)$ is maintained constant. Thus as time increases, say from t to $t + \Delta t$, we must move in the positive z direction a distance $\Delta z = v\,\Delta t$. Again, v is identified as the velocity of any part of the wave in the positive z direction.

Only one solution of the second degree differential equation has been given. A second solution may be written as any function of $t + (z/v)$

and checked by methods exactly similar to those used for the first solution. This is identified as a wave traveling in the negative z direction with velocity v. A complete solution to (1) is then

$$V = F_1\left(t - \frac{z}{v}\right) + F_2\left(t + \frac{z}{v}\right) \tag{4}$$

where

$$v = \frac{1}{\sqrt{LC}} \tag{5}$$

1.17 Relation between Voltage and Current in the Ideal Line

The solution of the differential Eq. 1.15(7) for voltage has been obtained as Eq. 1.16(4). Since current satisfies the same differential equation as does voltage, the solution for current must have the same form.

$$I = f_1\left(t - \frac{z}{v}\right) + f_2\left(t + \frac{z}{v}\right) \tag{1}$$

The first term in this solution also represents the function f_1 traveling with velocity v and unchanged shape in the positive z direction; the second term represents the function f_2 traveling in the negative z direction with velocity v and unchanged shape. The velocity is given in terms of the distributed constants of the line by Eq. 1.16(5).

By substituting (1) and Eq. 1.16(4) in Eq. 1.15(3),

$$-\frac{1}{v}F_1'\left(t - \frac{z}{v}\right) + \frac{1}{v}F_2'\left(t + \frac{z}{v}\right) = -L\left[f_1'\left(t - \frac{z}{v}\right) + f_2'\left(t + \frac{z}{v}\right)\right]$$

In the above, the primes indicate derivatives with respect to $t - (z/v)$. The positively traveling wave of current may be related directly to the positively traveling wave of voltage.

$$Lf_1'\left(t - \frac{z}{v}\right) = \frac{1}{v}F_1'\left(t - \frac{z}{v}\right) \tag{2}$$

or

$$f_1\left(t - \frac{z}{v}\right) = \frac{1}{vL}F_1\left(t - \frac{z}{v}\right) \tag{3}$$

Any constant term arising from the integration of (2) to obtain (3) is ignored since the present interest is only in wave phenomena.

Similarly, by relating the negatively traveling wave of current to the

negatively traveling wave of voltage

$$f_2\left(t+\frac{z}{v}\right) = -\frac{1}{vL}F_2\left(t+\frac{z}{v}\right) \tag{4}$$

so that (1) may be written

$$I = \frac{1}{Z_0}\left[F_1\left(t-\frac{z}{v}\right) - F_2\left(t+\frac{z}{v}\right)\right] \tag{5}$$

where vL is replaced by Z_0. By substituting the value of v from Eq. 1.16(5),

$$Z_0 = \sqrt{\frac{L}{C}} = vL = \frac{1}{vC} \tag{6}$$

The current in the positively traveling wave is then obtained by dividing the voltage of the positively traveling wave by the constant Z_0; current in the negatively traveling wave is the negative of the voltage of the negatively traveling wave divided by Z_0. Z_0 is the characteristic impedance or surge impedance of the line. For the ideal line it is a purely real quantity and is given in terms of the distributed constants of the line by any of the three forms of (6). We shall show in a later chapter that for an ideal line, the velocity v must be equal to the velocity of light in the dielectric material of the line.

1.18 Reflection and Transmission at a Discontinuity

Most transmission line problems are concerned with discontinuities. For example, a uniform transmission line of known characteristic impedance may be connected to another of different characteristic impedance, to a load impedance, or to some other type of discontinuity. We shall only assume for present purposes that at the point of the discontinuity, an impedance Z_2 can be calculated representing the ratio of voltage to current for the load, transmission line arrangement, or whatever else may be connected to a line of known characteristic impedance, Z_1 (Fig. 1.18).

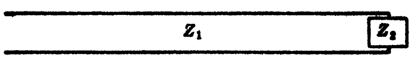


FIG. 1.18. Transmission line terminated in Z_2.

By Kirchhoff's laws, total voltage and current must be continuous across the discontinuity. The total voltage in the line may be regarded as the sum of voltage in a positively traveling wave, equal to V_1 at the point of the discontinuity, and voltage in a reflected or negatively traveling wave, equal to V_1' at the discontinuity. The sum of V_1 and V_1'

must be V_2, the voltage appearing across Z_2.

$$V_1 + V_1' = V_2 \qquad [1]$$

Similarly, the sum of currents in the positively and negatively traveling waves in the line, at the point of discontinuity, must be equal to the current I_2 flowing into Z_2.

$$I_1 + I_1' = I_2 \qquad [2]$$

But recall from Art. 1.17 that I_1 is obtained by dividing V_1 by Z_1, I_1' by dividing V_1' by $-Z_1$. Also, by definition of Z_2,

$$I_2 = \frac{V_2}{Z_2}$$

So (2) may be rewritten

$$\frac{V_1}{Z_1} - \frac{V_1'}{Z_1} = \frac{V_2}{Z_2} \qquad [3]$$

By combining (1) and (3), we may find the ratio of voltage in the negatively traveling wave to that in the positive,

$$\frac{V_1'}{V_1} = \frac{Z_2 - Z_1}{Z_2 + Z_1} = \frac{K - 1}{K + 1}$$

where

$$K = \frac{Z_2}{Z_1} \qquad [4]$$

This is the reflection coefficient in terms of voltage, since it is the voltage in the reflected wave compared to that in the incident wave.

Similarly from (1) and (3), we may find the ratio of voltage transmitted to Z_2 to that in the incident wave.

$$\frac{V_2}{V_1} = \frac{2Z_2}{Z_2 + Z_1} = \frac{2K}{K + 1} \qquad [5]$$

This is called the transmission coefficient in terms of voltages.

Similarly, by referring to the relations between currents and voltages,

$$\frac{I_1'}{I_1} = \frac{Z_1 - Z_2}{Z_1 + Z_2} = \frac{1 - K}{K + 1} \qquad [6]$$

$$\frac{I_2}{I_1} = \frac{2Z_1}{Z_2 + Z_1} = \frac{2}{K + 1} \qquad [7]$$

The most interesting, and probably the most obvious, conclusion from the above relations is this: there is no reflected wave if the termi-

nating impedance is exactly equal to the characteristic impedance of the line. All energy in the incident wave is then transferred to the impedance Z_2 which cannot be distinguished from a line of infinite length and characteristic impedance $Z_1 = Z_2$.

Other applications of these general relations to many types of problems will be found throughout the book. It might be emphasized here that the above relations give ratios to quantities in the incident or positively traveling wave, although any meters placed in the transmission line would measure total quantities, incident plus reflected.

1.19 Application of Traveling Wave Ideas to Some Simple Problems

(*a*) *Direct-current voltage applied to an infinite line.* Consider the case of a D-C voltage V, suddenly applied to an ideal line of infinite length (Fig. 1.19*a*). The line starts to charge to voltage V, the wave front

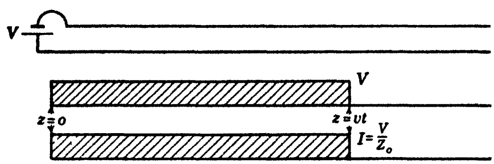


FIG. 1.19*a*. D–C voltage suddenly applied to an infinite line.

traveling with the velocity $v = 1/\sqrt{LC}$. Since there is never any discontinuity, there is never any reflected wave, and the only current is that flowing in the positive wave, V/Z_0. This then is a D-C current flowing to the charges which appear on the line as voltage moves along. At any time t after impressing the voltage, there is voltage V and current V/Z_0 in the line up to the point $z = vt$, and no voltage or current beyond.

(*b*) Suppose now that the infinite line of part (*a*) is suddenly connected to a D-C voltage V, the voltage now being applied not at one end but rather at the center of the line. The part of the line to the right of the center must experience the same charging caused by the positive traveling wave as was described in (*a*). The line to the left must, from symmetry, have exactly the same experience as the line to the right, except, of course, that its wave will be a negative traveling wave since the direction to the right has been taken as positive. The voltages on each side, at the same distance from the source, must be identical at every instant. The currents must be equal but opposite at corresponding points since positive current is taken as current flowing in the positive direction. This checks with the voltage and current relations of Art. 1.17 where with the positively traveling wave's current taken as

V/Z_0, the negatively traveling wave's current naturally comes out to be $-V/Z_0$.

(c) *Direct-current voltage applied to a shorted line.* Suppose that the D-C voltage is applied to a line which is not infinite in length, but is shorted at some point, $z = l$ (Fig. 1.19b). We know that finally infinite current will flow if V is maintained. However, the mechanism of current build-up is at least interesting. After voltage is applied to the line,

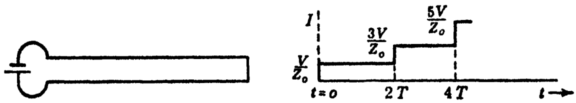


FIG. 1.19b. D-C voltage suddenly applied to a shorted line.

everything proceeds as in (a) until the time that the wave reaches the short circuit. At the time the incident wave with voltage V appears across the short circuit, which demands zero voltage, a reflected or negatively traveling wave is sent back of voltage $(-V)$, so that the sum of voltages in the two waves is indeed zero. Since current in the negative traveling wave is the negative of voltage divided by Z_0, this is $-(-V/Z_0)$ or $+V/Z_0$ and so adds directly to the current in the positive traveling wave. This reflected wave then moves to the left, leaving a wake of zero voltage and a current equal to $2V/Z_0$ behind it. As soon as the reflected wave has traveled back to the source, it brings the zero voltage condition back to this point so that the D-C voltage must send out a new wave of voltage V down the line, with associated current V/Z_0, making a total current in the line $3V/Z_0$ at this time. Current then builds up to infinity in the step manner indicated by Fig. 1.19b. The time T is the time for a wave to travel one way down the line,

$$T = \frac{l}{v}$$

(d) *Shorting of a charged line.* Consider a transmission line open on both ends and charged to a D-C potential, V. If one end is shorted,

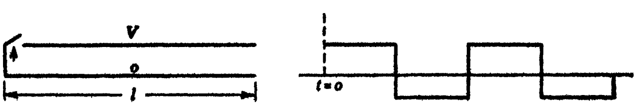


FIG. 1.19c. Charged line of length l suddenly shorted.

as by the switch in Fig. 1.19c, a wave of voltage $(-V)$ must be started down the line, since this wave must add to the D-C condition V to give

zero at the short circuit. The current connected with the wave is then $-V/Z_0$. When this wave reaches the open end, there must then be started a reflected wave such that total current at this open end is zero. Current in this reflected wave must then be $+V/Z_0$. Because current in a negatively traveling wave is the negative of voltage divided by Z_0, this will require a voltage $-V$ for the reflected wave. Thus the wave traveling to the right wipes out the original D-C voltage; that traveling back causes it to appear on the other side. The current required for the interchange of charge flows through the short during the entire time.

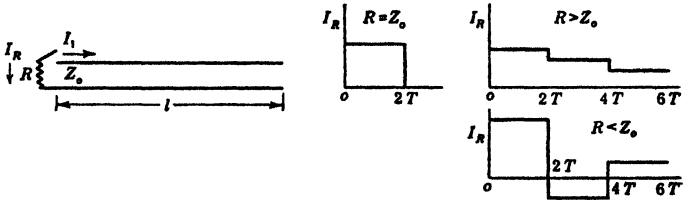


Fig. 1.19d. Charged line of length l suddenly connected to a resistor.

As soon as the reflected wave arrives back at the source, the line is charged up in the opposite direction with no current flowing in the line, so the process may repeat in the opposite direction. Current and voltage relations are indicated in Fig. 1.19c. It is evident that this is an oscillating system. The problem will be considered by another method in a later section.

(*e*) *Charged line connected to a resistor.* If the charged line of (*c*) is connected to a resistor instead of a short circuit, the amount of wave started down the line may be determined as follows.

The voltage across the resistance is the sum of the D-C voltage of the line and the voltage in the wave, V_1.

$$V_R = V + V_1 \qquad [1]$$

The current flowing into the resistor is merely the negative of current for the positively traveling wave.

$$I_R = -I_1$$

or

$$\frac{V_R}{R} = -\frac{V_1}{Z_0} \qquad [2]$$

By combining (1) and (2)

$$V_R = V\left(\frac{R}{R+Z_0}\right) = -V_1\frac{R}{Z_0} \qquad [3]$$

For example, if $R = Z_0$, the voltage appearing across the resistance at the first instant is half the D-C voltage of the line, as is the voltage appearing in the traveling wave. When this wave reaches the open end of the line, the current zero requirement produces a reflected wave whose voltage is equal to and in the same direction as that in the incident wave, as found by the reasoning of part (d). Thus in the case of $R = Z_0$, the original wave wipes out half the voltage, and the corresponding current, $-V/2Z_0$, is that which flows through R. The reflected wave wipes out the remaining half of the voltage and, of course, reduces current to zero. When this wave reaches $R = Z_0$ there is no further reflection, so all is still. Current wave shape is shown in Fig. 1.19d. Also shown are currents for $R > Z_0$ and $R < Z_0$.

1.20 Approximate Attenuation in Lines with Losses

All previous results have applied to ideal transmission lines. If losses are present but small, as in many practical cases at high frequencies, the first correction due to these losses may be approximated. The method is similar to that used in finding the exponential decay with time for natural oscillations in a lossy circuit. This method will also be invaluable for later analyses of wave guides. It assumes that the behavior is given primarily by the solution obtained with losses completely neglected; and that the main correction is obtained by taking currents and voltages the same as in the ideal case, but allowing them to encounter the known resistances and conductances.

The major correction due to losses will appear as an attenuation down the line. Assuming that this attenuation is exponential of the form $e^{-\alpha z}$, the previous solutions for voltage and current may be multiplied by such exponentials. Consider only a positively traveling wave.

$$V = e^{-\alpha z} F\left(t - \frac{z}{v}\right) \tag{1}$$

and

$$I = \frac{1}{Z_0} e^{-\alpha z} F\left(t - \frac{z}{v}\right) \tag{2}$$

The power transferred down the line by this wave at any point and any instant is merely the product of V and I.

$$W_T = VI = \frac{e^{-2\alpha z}}{Z_0}\left[F\left(t - \frac{z}{v}\right)\right]^2 \text{ watts} \tag{3}$$

The time average power transfer across any point along the line is found by averaging $\left[F\left(t - \frac{z}{v}\right)\right]^2$. Let the average of this be K. Then

$$(W_T)_{\text{av}} = \frac{Ke^{-2\alpha z}}{Z_0} \tag{4}$$

The rate of decrease of this average power transfer with distance along the line must correspond to the average power loss in the line per unit length.

$$(W_L)_{\text{av}} = -\frac{d(W_T)_{\text{av}}}{dz} = 2\alpha\,\frac{Ke^{-2\alpha z}}{Z_0} = 2\alpha(W_T)_{\text{av}}$$

or

$$\alpha = \frac{(W_L)_{\text{av}}}{2(W_T)_{\text{av}}} \tag{5}$$

To the extent of the approximations inherent in the above analyses, the attenuation factor α is then given by the average power loss per unit length divided by twice the average power transferred down the line.

If the current (2) flows through a resistance R per unit length, it produces a loss

$$W_R = I^2R = \frac{e^{-2\alpha z}}{Z_0^2}\left[F\left(t - \frac{z}{v}\right)\right]^2 R$$

Since K is defined above as the average of $\left[F\left(t - \frac{z}{v}\right)\right]^2$, the average loss in the resistance is

$$(W_R)_{\text{av}} = \frac{RK}{Z_0^2}\,e^{-2\alpha z} \text{ watts/unit length} \tag{6}$$

If the voltage of the line, (1), appears across an imperfect dielectric such that there is a conductance G per unit length, the loss produced is

$$W_G = V^2G = Ge^{-2\alpha z}\left[F\left(t - \frac{z}{v}\right)\right]^2$$

or

$$(W_G)_{\text{av}} = KGe^{-2\alpha z} \text{ watts/unit length} \tag{7}$$

The total loss per unit length is the sum of (6) and (7).

$$(W_L)_{\text{av}} = Ke^{-2\alpha z}\left[\frac{R}{Z_0^2} + G\right] \tag{8}$$

If (4) and (8) are substituted in (5),

$$\alpha = \frac{R}{2Z_0} + \frac{GZ_0}{2} \qquad [9]$$

The attenuation is obtained by this approximate analysis in terms of the constants of the line. All other properties of the line (characteristic impedance, velocity of propagation, etc.) are assumed to be given well enough by results of the previous analyses of the loss-free case.

Units of α as given by (9) are in nepers per unit length. The particular unit length to use is determined by that used to measure R and G. To convert nepers to decibels, multiply by 8.686.

1.21 Ideal Line with Applied Sinusoidal Voltages

Much of the preceding discussion has involved little restriction on the type of variation with time of the voltages applied to the transmission lines. Most practical problems are concerned entirely or at least partially with sinusoidal time variations. If a voltage which is sinusoidal in time is applied at $z = 0$, it may be represented by the exponential (see Art. 1.11)

$$V\big|_{z=0} = F(t) = V_1 e^{j\omega t} \qquad [1]$$

then the corresponding positively traveling wave is written

$$V_1 e^{j\omega\left(t - \frac{z}{v}\right)}$$

Similarly, a negatively traveling wave is written

$$V_1' e^{j\omega\left(t + \frac{z}{v}\right)}$$

Or the total solution, made up of positive and negative traveling waves,

$$V = e^{j\omega t}\left[V_1 e^{-j\frac{\omega z}{v}} + V_1' e^{j\frac{\omega z}{v}}\right] \qquad [2]$$

The corresponding current, from Art. 1.17, is

$$I = \frac{e^{j\omega t}}{Z_0}\left[V_1 e^{-j\frac{\omega z}{v}} - V_1' e^{j\frac{\omega z}{v}}\right] \qquad [3]$$

For problems in which we shall be concerned throughout with sinusoidal quantities, it is not necessary to write the factor $e^{j\omega t}$ explicitly each time, since it will always be understood that all terms are multiplied by this factor. We rewrite (2) and (3), omitting it.

$$V = V_1 e^{-j\beta z} + V_1' e^{j\beta z} \qquad [4]$$

$$I = \frac{1}{Z_0}\left[V_1 e^{-j\beta z} - V_1' e^{j\beta z}\right] \qquad [5]$$

where

$$\beta = \frac{\omega}{v} \qquad [6]$$

The quantity β is called the phase constant of the line, logically so, since βz measures the phase angle of a voltage or current in a single wave for any point z, with respect to voltage or current at that same instant at $z = 0$. Moreover, if voltage and current are observed at any point z, they will be found exactly the same at points such that βz differs from that of the first point by multiples of 2π. The distance between points of like current and voltage is called a wavelength, λ. By the above reasoning,

$$\beta\lambda = 2\pi$$

or

$$\beta = \frac{2\pi}{\lambda} \qquad [7]$$

Finally, the velocity v may be termed a phase velocity since it is the velocity with which a point of constant phase (total phase, $\omega t - \beta z$) moves. That is, to maintain

$$\omega t - \beta z = \text{Constant}$$

$$\frac{dz}{dt} = \frac{\omega}{\beta} = v$$

from (6). More will be said about phase velocity later.

Problem 1.21 (a). Show that the input impedance of an ideal transmission line of characteristic impedance Z_0 and length l terminated in an output impedance Z_L is

$$Z_i = Z_0 \frac{Z_L \cos \beta l + jZ_0 \sin \beta l}{Z_0 \cos \beta l + jZ_L \sin \beta l}$$

Problem 1.21 (b). When two transmission lines are to be connected in cascade, a reflection of the wave to be transmitted from one to the other will occur if they do

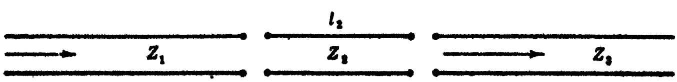


FIG. 1.21. Matching section for matching Z_1 to Z_3.

not have the same characteristic impedances. Show from the relations of Art. 1.21 and Problem 1.21 (a) that a quarter-wave line matching transformer (Fig. 1.21) will cause the first line to see its characteristic impedance Z_1 as a termination and thus eliminate reflection in transfer if $\beta l_2 = \pi/2$ and $Z_2 = \sqrt{Z_1 Z_3}$.

1.22 Shorted Lines; Standing Waves

Suppose a transmission line, shorted at one end, is excited by sinusoidal voltage at the other. Let us select the position of the short as the reference, $z = 0$. Voltage and current at any point along the line may be written as the sum of an incident and a reflected wave, as in Eqs. 1.21(4) and 1.21(5). The short imposes the condition that at $z = 0$, voltage must always be zero. From Eq. 1.21(4),

$$V\Big|_{z=0} = V_1 + V_1'$$

For this to be zero, V_1' must be the negative of V_1. This result could be obtained as well from the general results for reflections at a discontinuity by setting $Z_2 = 0$ in Eq. 1.18(4), or merely by physical reasoning which shows that no energy is absorbed by the short circuit, so all energy brought by the incident wave must appear in the reflected wave. The two waves of equal energy in the same line must have equal voltages. These must be in opposite directions at the short to add to the required zero voltage.

If $V_1' = -V_1$ is substituted in Eqs. 1.21(4) and 1.21(5),

$$V = V_1[e^{-j\beta z} - e^{j\beta z}] = -2jV_1 \sin \beta z \qquad [1]$$

$$I = \frac{V_1}{Z_0}[e^{-j\beta z} + e^{j\beta z}] = 2\frac{V_1}{Z_0} \cos \beta z \qquad [2]$$

The above results, typical for standing waves, show the following.

1. Voltage is always zero not only at the short, but also at multiples of $\lambda/2$ to the left. That is,

$$V = 0 \text{ at } -\beta z = n\pi \quad \text{or} \quad z = -n\frac{\lambda}{2}$$

2. Voltage is a maximum at all points for which βz is an odd multiple of $\pi/2$. These are at distances odd multiples of a quarter wavelength from the short circuit.

$$V = \text{maximum at } -\beta z = \frac{(2m + 1)\pi}{2} \quad \text{or} \quad z = -\frac{(2m+1)\lambda}{4}$$

3. Current is a maximum at the short circuit and at all points where voltage is zero; it is zero at all points where voltage is a maximum.

4. Current and voltage are not only displaced in their space patterns, but also are 90° out of time phase, as indicated by the j appearing in (1).

5. The ratio between the maximum current on the line and the maximum voltage is Z_0, the characteristic impedance of the line.

6. The total energy in any length of line a multiple of a quarter wavelength long is constant, merely interchanging between energy in the electric field of the voltages and energy in the magnetic field of the currents.

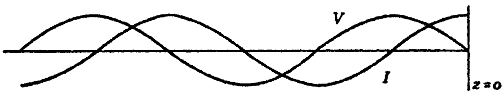


FIG. 1.22. Standing waves of voltage and current along shorted line.

To check the energy relation stated above, calculate the magnetic energy of the currents at a time when the current pattern is a maximum and voltage is zero everywhere along the line. Current is given by (2). The energy is calculated for a quarter wavelength of the line.

$$U_M = \frac{L}{2}\int_{-\lambda/4}^{0} |I|^2\,dz = \frac{L}{2}\int_{-\lambda/4}^{0} \frac{4V_1^2}{Z_0^2}\cos^2\beta z\,dz$$

$$= \frac{2V_1^2L}{Z_0^2}\left[\frac{z}{2} + \frac{1}{4\beta}\sin 2\beta z\right]_{-\lambda/4}^{0}$$

Since $\beta = 2\pi/\lambda$ by 1.21(7), the above is simply

$$U_M = \frac{V_1^2L\lambda}{4Z_0^2} \qquad [3]$$

The maximum energy stored in the distributed capacity effect of the line is calculated for the quarter wavelength when the voltage pattern is a maximum and current is everywhere zero. Voltage is given by (1).

$$U_E = \frac{C}{2}\int_{-\lambda/4}^{0} |V|^2\,dz = \frac{C}{2}\int_{-\lambda/4}^{0} 4V_1^2\sin^2\beta z\,dz$$

$$= 2CV_1^2\left[\frac{z}{2} - \frac{1}{4\beta}\sin^2\beta z\right]_{-\lambda/4}^{0} = \frac{CV_1^2\lambda}{4} \qquad [4]$$

By the definition of Z_0, (3) may also be written

$$U_M = \frac{V_1^2L\lambda}{4\dfrac{L}{C}} = \frac{V_1^2C\lambda}{4} = U_E \qquad [5]$$

Thus the maximum energy stored in magnetic fields is exactly equal to that stored in electric fields 90° later in time. It could actually be shown that the sum of electric and magnetic energy at any other part of the cycle is equal to this same value. This suggests a considerable

amount of similarity with the parallel resonant circuit treated in Arts. 1.08, 1.09, and 1.10. The similarity may be used to yield some very convenient ways of looking at shorted lines as the equivalent of tuned circuits for many practical problems. For example, whenever the line is shorted at a distance of

$$l = \frac{(2n - 1)\lambda}{4}; \quad n = 1, 2, 3 \cdots \qquad [8]$$

from the source, the input impedance becomes very high. The source is then located at a current node and a voltage loop and the line is spoken of as resonant. When losses are considered it will be found that comparisons of stored energy with energy dissipated per radian will lead to a value of Q which can be used as a means of gaging the variation of impedance near resonance just as in simple circuits.

Also, given any section of line, shorted on both ends, there are an infinite number of natural frequencies for which, as required by the boundary conditions, the voltage nodes will fall on the ends of the line. This will occur whenever

$$l = \frac{n\lambda}{2} \qquad [9]$$

Problem 1.22. Find the Q and the input impedance of a shorted quarter wavelength line, using the approximate loss formulas of Art. 1.20.

1.23 Combinations of Natural Modes to Fit Initial Conditions

In this section we shall demonstrate a technique which will be one of the most widely useful methods for solution of field and wave problems to come later. The method makes use of a summation or series of harmonic solutions to a wave problem to fit imposed boundary or initial conditions, just as in Art. 1.13 a series of sinusoids was used to fit any arbitrary periodic functions.

As the example, let us consider a problem quite similar to that solved by a straightforward traveling wave analysis in Art. 1.19*d*. For this problem, imagine the open-circuited transmission line, first charged to a D-C voltage V_0, and then shorted at both ends simultaneously at a specified instant of time. The voltage distribution at the instant of shorting is then known (zero at each end and a constant equal to V_0 at all other points, as sketched in Fig. 1.23*a*). It is desired to find the current and voltage behavior at all later times.

In Art. 1.22 it was noted that natural sinusoidal oscillations for a line of length l, shorted at both ends, occur at all frequencies for which the line is a multiple of a half-wave long. From the results of Eq. 1.22(9),

one of these natural sinusoidal modes of oscillation may be written. For voltage,

$$V_m = A_m e^{j\omega_m t} \sin \frac{m\pi z}{l} \qquad [1]$$

where

$$\omega_m = 2\pi f_m = \frac{m\pi v}{l} = \frac{m\pi}{\tau} \qquad [2]$$

The time of travel of a wave down the line is $\tau = l/v$. The results of Art. 1.22 also reveal that the corresponding current is 90° out of time and space phase with voltage and has the magnitude of voltage divided by Z_0. That is,

$$I_m = \frac{jA_m}{Z_0} e^{j\omega_m t} \cos \frac{m\pi z}{l} \qquad [3]$$

Now, let us form a solution to the transmission line equations from the sum of all solutions of the form of (1). The basis for this step may be traced to the fact that the sum of solutions to a linear differential equation is also a solution. The transmission line differential equations are linear, and (1) is a solution. Adding,

$$V = A_1 e^{j\omega_1 t} \sin \frac{\pi z}{l} + A_2 e^{j\omega_2 t} \sin \frac{2\pi z}{l} + A_3 e^{j\omega_3 t} \sin \frac{3\pi z}{l} + \cdots \qquad [4]$$

and the corresponding sum of (3) for current,

$$I = \frac{j}{Z_0}\left[A_1 e^{j\omega_1 t} \cos \frac{\pi z}{l} + A_2 e^{j\omega_2 t} \cos \frac{2\pi z}{l} + A_3 e^{j\omega_3 t} \cos \frac{3\pi z}{l} + \cdots\right] \qquad [5]$$

The amplitudes $A_1, A_2, \cdots A_m$ are still arbitrary. They may be determined from the known initial condition by expanding the known initial voltage distribution with distance as a Fourier series.

The previous introduction of the Fourier series was for use with periodic time functions. Certainly its use is not restricted to time as its variable (see Prob. 1.14*c*), for a periodic function of any variable (for example, distance) may be expanded in a similar series of sinusoids. The usefulness for functions which are to be represented over a certain limited range, although these functions are not necessarily periodic, is also well known. For example, if we wish to represent the rectangular function of distance shown in Fig. 1.23*a*, we require only that this representation be accurate over the limited range 0 to l. This rectangle might then be considered as one rectangle from a repeating periodic function such as that of time sketched in Fig. 1.14. The fact that the series

actually represents a function that repeats indefinitely causes no worry since we do not care what happens outside the range 0 to l.

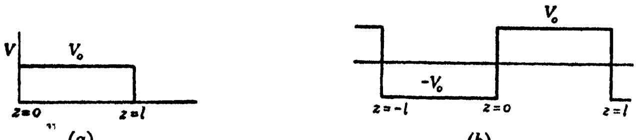


FIG. 1.23.

The results of changing previous Fourier coefficients for time functions to a form suitable for use with z, taking $2l$ as the period, are

$$a_0 = \frac{1}{2l}\int_{-l}^{+l} f(z)\,dz \tag{6}$$

$$a_n = \frac{1}{l}\int_{-l}^{+l} f(z)\cos\frac{n\pi z}{l}\,dz \tag{7}$$

$$b_n = \frac{1}{l}\int_{-l}^{+l} f(z)\sin\frac{n\pi z}{l}\,dz \tag{8}$$

Since it is immaterial what happens to the function outside the bounds of the line, let us write a series for the odd function of Fig. 1.23b, which at least represents the initial voltage distribution over the length 0 to l. This will then have sine terms only (see Prob. 1.14b), and

$$b_n = \frac{2}{l}\int_0^{l} f(z)\sin\frac{n\pi z}{l}\,dz \tag{9}$$

where

$$f(z) = V_0, \quad 0 < z < l \quad \text{and} \quad f(z) = 0 \text{ at } z = 0,\ z = l$$

$$b_n = \frac{2}{l}\int_0^{l} V_0 \sin\frac{n\pi z}{l}\,dz = -\frac{2V_0}{l}\cdot\frac{l}{n\pi}\cos\frac{n\pi z}{l}\Big|_0^l$$

$$b_n = \frac{2V_0}{n\pi}[1 - \cos n\pi]$$

Since $\cos n\pi$ is -1 if n is odd, and $+1$ if n is even,

$$b_n = 0,\ n \text{ even}$$

$$b_n = \frac{4V_0}{n\pi},\ n \text{ odd}$$

Thus the Fourier series representing the initial voltage distribution over the line may be written

$$V\Big|_{t=0} = \frac{4V_0}{\pi}\left[\sin\frac{\pi z}{l} + \frac{1}{3}\sin\frac{3\pi z}{l} + \frac{1}{5}\sin\frac{5\pi z}{l} + \cdots\right] \qquad [10]$$

But at $t = 0$, the series (4) reduces to

$$V\Big|_{t=0} = A_1 \sin\frac{\pi z}{l} + A_2 \sin\frac{2\pi z}{l} + A_3 \sin\frac{3\pi z}{l} + \cdots \qquad [11]$$

By a term-by-term comparison between (10) and (11) it is seen that all the even coefficients, A_2, A_4, etc., must be zero. For the odd coefficients,

$$A_1 = \frac{4V_0}{\pi}, \quad A_3 = \frac{4V_0}{3\pi}, \quad \cdots A_n = \frac{4V_0}{n\pi}$$

Now that the coefficients are determined, the complete series expressions for voltage and current at any time may be written

$$V = \frac{4V_0}{\pi}\left[e^{j\frac{\pi t}{\tau}}\sin\frac{\pi z}{l} + \frac{e^{j\frac{3\pi t}{\tau}}}{3}\sin\frac{3\pi z}{l} + \frac{e^{j\frac{5\pi t}{\tau}}}{5}\sin\frac{5\pi z}{l} + \cdots\right] \qquad [12]$$

$$I = \frac{4jV_0}{Z_0\pi}\left[e^{j\frac{\pi t}{\tau}}\cos\frac{\pi z}{l} + \frac{e^{j\frac{3\pi t}{\tau}}}{3}\cos\frac{3\pi z}{l} + \frac{e^{j\frac{5\pi t}{\tau}}}{5}\cos\frac{5\pi z}{l} + \cdots\right] \qquad [13]$$

These series forms may be used to calculate current or voltage at any point z along the line at any time t after closing the shorting switches. The desired accuracy determines the number of terms that must be retained. It is especially interesting to note the current through the short circuit at the end, by letting $z = 0$ in (13),

$$I\Big|_{z=0} = \frac{4jV_0}{Z_0\pi}\left[je^{j\frac{\pi t}{\tau}} + j\frac{e^{j\frac{3\pi t}{\tau}}}{3} + j\frac{e^{j\frac{5\pi t}{\tau}}}{5} + \cdots\right] \qquad [14]$$

Recalling that $je^{j\frac{\pi t}{\tau}}$ is a representation for $\sin \pi t/\tau$ (Art. 1.11), and, comparing with the Fourier analyses for square waves in this article and in Art. 1.14, it is found that (14) is the Fourier series for such a square wave of current as a function of time. The result for current, if examined carefully, is then found to be exactly the same as that obtained by a traveling wave reasoning for the similar example of Art. 1.22.

The value of this method cannot be fully appreciated by a single example. There will be many other examples in later chapters in which a series of separate solutions will be formed to fit known boundary or initial conditions.

1.24 Transmission Lines with Losses. Solution by Differential Equations for Sinusoidal Voltages

We have obtained an attenuation factor for the line by an approximate method in Art. 1.20. Let us now examine the effect of a series resistance and a shunt conductance more carefully by inserting the effects directly into the differential equations of the line.

If series resistance and shunt conductance are of importance in the transmission lines, the voltage drop along the line must include the resistance drop as well as the inductance drop of Eq. 1.15(3). Similarly, the leakage current must include the conductance as well as capacitance current shunted across the line. Instead of Eqs. 1.15(3) and 1.15(4) we then have

$$\frac{\partial V}{\partial z} = -L\frac{\partial I}{\partial t} - RI \tag{1}$$

$$\frac{\partial I}{\partial z} = -C\frac{\partial V}{\partial t} - GV \tag{2}$$

If steady state sinusoidal conditions with respect to time are considered, of the form $e^{j\omega t}$, time derivatives may be replaced by $j\omega$, and total derivatives written for distance since there are then no other derivatives.

$$\frac{dV}{dz} = -(R + j\omega L)I \tag{3}$$

$$\frac{dI}{dz} = -(G + j\omega C)V \tag{4}$$

Differentiate (3) with respect to z and substitute (4).

$$\frac{d^2V}{dz^2} = -(R + j\omega L)\frac{dI}{dz} = (R + j\omega L)(G + j\omega C)V$$

or

$$\frac{d^2V}{dz^2} = \gamma^2 V \tag{5}$$

where

$$\gamma^2 \equiv (R + j\omega L)(G + j\omega C) \tag{6}$$

The solution to (5) is in terms of exponentials,

$$V = Ae^{-\gamma z} + Be^{+\gamma z} \tag{7}$$

as can be verified by substituting (7) in (5).

An examination of (6) shows that γ must be complex in the general case. Let us write the real part as α, the imaginary part as β. By (6),

$$\gamma = \alpha + j\beta = \sqrt{(R + j\omega L)(G + j\omega C)} \tag{8}$$

If (7) is rewritten using α and β,

$$V = Ae^{-\alpha z}e^{-j\beta z} + Be^{\alpha z}e^{j\beta z} \tag{9}$$

This expression for voltage is quite similar to that of Eq. 1.21(4) for sinusoidal waves in ideal lines, except that there is now an attenuation term on both the forward and backward traveling waves. α is called the attenuation constant, β the phase constant, and γ the propagation constant. The relation between these results and those of previous expansions is best seen by a binomial expansion of (8), valid when $R/\omega L \ll 1$ and $G/\omega C \ll 1$. Then

$$\{[R + j\omega L][G + j\omega C]\}^{1/2} = \left[\frac{R}{2\sqrt{\frac{L}{C}}} + \frac{\sqrt{\frac{L}{C}}G_0}{2}\right] + j\omega\sqrt{LC}\left[1 - \frac{RG}{4\omega^2 LC} + \frac{G^2}{8\omega^2 C^2} + \frac{R^2}{8\omega^2 L^2}\right]$$

or

$$\alpha \cong \frac{R}{2\sqrt{\frac{L}{C}}} + \frac{G_0\sqrt{\frac{L}{C}}}{2} \tag{10}$$

$$\beta \cong \omega\sqrt{LC}\left[1 - \frac{RG}{4\omega^2 LC} + \frac{G^2}{8\omega^2 C^2} + \frac{R^2}{8\omega^2 L^2}\right] \tag{11}$$

Equation (10) is seen to be of exactly the same form as Eq. 1.20(9). Equation (11) shows that β is so little different from the value $\omega\sqrt{LC}$ [Eq. 1.21(6)], at least for low-loss lines, that it is usually sufficiently accurate to use this value.

$$\beta = \frac{\omega}{v} \cong \omega\sqrt{LC} = \frac{2\pi}{\lambda} \tag{12}$$

If the above approximations are not sufficiently good, it is possible to calculate more accurate values by obtaining the real and imaginary parts of (8).

As for current, solve by substituting (7) in (3)

$$I = \frac{\gamma}{R + j\omega L}\left[Ae^{-\gamma z} - Be^{\gamma z}\right]$$

or

$$I = \frac{1}{Z_0}\left[Ae^{-\alpha z}e^{-j\beta z} - Be^{\alpha z}e^{j\beta z}\right] \qquad [13]$$

where

$$Z_0 = \frac{R + j\omega L}{\gamma} = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \qquad [14]$$

Z_0 may again be thought of as the characteristic or surge impedance of the line, since it relates voltage and current in a single wave and is the impedance of a line of infinite length. However, it is now complex. Equation (14) may also be subjected to a binomial expansion and higher order terms neglected, subject to the condition that $R/\omega L \ll 1$ and $G/\omega C \ll 1$. Then

$$Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}}$$

$$\cong \sqrt{\frac{L}{C}}\left[\left(1 + \frac{R^2}{8\omega^2 L^2} - \frac{3G^2}{8\omega^2 C^2} + \frac{RG}{4\omega^2 LC}\right) + j\left(\frac{G}{2\omega C} - \frac{R}{2\omega L}\right)\right] \qquad [15]$$

The major correction is the reactance term which now appears in Z_0. However, for many practical lines it is sufficiently good to neglect all corrections and use only

$$Z_0 \cong \sqrt{\frac{L}{C}} \qquad [16]$$

When this is not accurate enough, approximate corrections may be added by (15), or the value calculated from the complete expression (14).

1.25 Velocities of Wave Propagation

In Art. 1.16 it was shown that for a perfectly conducting transmission line, a voltage applied to the line at one point will appear later in time, reproduced exactly in wave shape, along the line at some distance from the source. This conclusion was obtained from the form of the solution to the differential equation of the line. A certain velocity v, a quantity appearing in the original differential equation

$$\frac{\partial^2 V}{\partial z^2} = \frac{1}{v^2}\frac{\partial^2 V}{\partial t^2}$$

was taken as the velocity of the wave. This again came naturally out of the solution, once it was decided to label it as a traveling wave.

Now this velocity v is, for perfect conductors, entirely independent of the shape of the wave. It is the velocity of the wave front or peak, or of any other point of the wave which makes a convenient marker. For a steady state, sinusoidally varying voltage with time, it may be a bit difficult to conceive of a ready means of putting a marker on one loop and following it down the line, distinguishing it from those before and after it. But still, the velocity v is the velocity with which an imaginary observer would have to move to see always the same instantaneous value of voltage or current. Since this says, in effect, that the observer maintains the same phase angle along the infinite stretch of sine wave, the velocity was called a phase velocity.

The situation described is not always true in a wave guiding system. In a transmission line with dissipation and in general electromagnetic wave guiding systems, there will in general be a change of wave shape in a complex wave put into the system by the source. This is explained and analyzed by thinking of any impressed voltage or current as made up of a series of sinusoidally varying quantities of such frequencies and amplitudes as are necessary to depict the true wave form faithfully. Then, as this group of sine waves progresses down the line, there will in general be a difference in the various phase velocities of the various sine waves. They will change their relative positions ("faster" waves will speed ahead, "slower" ones fall back) and, as in a dissipative line, change their amplitudes in a way which varies with their frequencies. The resultant wave at some distant point along the line may be very different in appearance from that which went in.

It thus becomes difficult always to decide just what is to be meant by wave-front velocity or any velocity associated with some marker on the true complex wave. The phase velocity still is a readily applied concept as before. It still is associated with steady state sine waves and so is a useful thing in analysis. But it should be recognized that the phase velocity will in general vary with frequency.

Often in radio problems, the wave being transmitted consists of a bundle of sine waves covering a frequency band which is small compared with the average frequency. An example is the two-term combination:

$$\sin (\omega_0 - \Delta\omega)t + \sin (\omega_0 + \Delta\omega)t \qquad [1]$$

If the above represents the transmitted voltage, then the voltage everywhere along the path (assuming no amplitude change) is

$$\sin [(\omega_0 - \Delta\omega)t - (\beta_0 - \Delta\beta)x] + \sin [(\omega_0 + \Delta\omega)t - (\beta_0 + \Delta\beta)x] \quad [2]$$

in which β is to be regarded as a function of frequency as indicated by the use of $\Delta\beta$ to go with $\Delta\omega$.

Expression (2) may be changed to

$$2 \cos (\Delta\omega t - \Delta\beta x) \sin (\omega t - \beta x)$$

which shows that the resultant voltage on the line at any point may be pictured as a high-frequency wave whose amplitude varies at a low-frequency rate. The envelope of the wave, in other words. is

$$\cos (\Delta\omega t - \Delta\beta x) \qquad [3]$$

It varies sinusoidally with both time and distance and thus may be regarded as a traveling wave. It is readily seen that the velocity of an imaginary observer who stays on the same point of the *envelope* is

$$v_g = \frac{d\omega}{d\beta} \qquad [4]$$

This is called the group velocity. Since the phase velocity [Eq. 1.24(12)] is

$$v_p = \frac{\omega}{\beta} \qquad [5]$$

the group velocity is seen to be

$$v_g = \frac{v_p}{1 - \dfrac{\omega}{v_p}\dfrac{dv_p}{d\omega}} \qquad [6]$$

1.26 Summary of Uniform Line Equations

Transmission line information in this chapter has been included for the specific purpose of presenting a wave point of view for later use. There was consequently no attempt at completeness in the information for one who is interested primarily in transmission line results; but to increase the usefulness of the book in this respect the most often used of the equations are summarized in Table 1.26. It should be noted, however, that for detailed design of transmission lines, circle diagrams and transmission line charts often provide the most convenient forms for using this information.

TABLE 1.26

Quantity	General Line	Ideal Line	Approximate Results for Low-Loss Lines
Propagation constant $\gamma = \alpha + j\beta$	$\sqrt{(R + j\omega L)(G + j\omega C)}$	$j\omega\sqrt{LC}$	(See α and β below)
Phase constant β	$Im(\gamma)$	$\omega\sqrt{LC} = \frac{\omega}{v} = \frac{2\pi}{\lambda}$	$\omega\sqrt{LC}\left[1 - \frac{RG}{4\omega^2 LC} + \frac{G^2}{8\omega^2 C^2} + \frac{R^2}{8\omega^2 L^2}\right]$
Attenuation constant α	$Re(\gamma)$	0	$\frac{R}{2Z_0} + \frac{GZ_0}{2}$
Characteristic impedance Z_0	$\sqrt{\frac{R + j\omega L}{G + j\omega C}}$	$\sqrt{\frac{L}{C}}$	$\sqrt{\frac{L}{C}}\left[1 + j\left(\frac{G}{2\omega C} - \frac{R}{2\omega L}\right)\right]$
Input impedance Z_i	$Z_0\left[\frac{Z_L \cosh \gamma l + Z_0 \sinh \gamma l}{Z_0 \cosh \gamma l + Z_L \sinh \gamma l}\right]$	$Z_0\left[\frac{Z_L \cos \beta l + jZ_0 \sin \beta l}{Z_0 \cos \beta l + jZ_L \sin \beta l}\right]$	
Impedance of shorted line	$Z_0 \tanh \gamma l$	$jZ_0 \tan \beta l$	$Z_0\left[\frac{\alpha l \cos \beta l + j \sin \beta l}{\cos \beta l + j\alpha l \sin \beta l}\right]$
Impedance of open line	$Z_0 \coth \gamma l$	$-jZ_0 \cot \beta l$	$Z_0\left[\frac{\cos \beta l + j\alpha l \sin \beta l}{\alpha l \cos \beta l + j \sin \beta l}\right]$
Impedance of quarter-wave line	$Z_0\left[\frac{Z_L \sinh \alpha l + Z_0 \cosh \alpha l}{Z_0 \sinh \alpha l + Z_L \cosh \alpha l}\right]$	$\frac{Z_0^2}{Z_L}$	$Z_0\left[\frac{Z_0 + Z_L \alpha l}{Z_L + Z_0 \alpha l}\right]$
Impedance of half-wave line	$Z_0\left[\frac{Z_L \cosh \alpha l + Z_0 \sinh \alpha l}{Z_0 \cosh \alpha l + Z_L \sinh \alpha l}\right]$	Z_L	$Z_0\left[\frac{Z_L + Z_0 \alpha l}{Z_0 + Z_L \alpha l}\right]$
Voltage along line $V(z)$	$V_i \cosh \gamma z - I_i Z_0 \sinh \gamma z$	$V_i \cos \beta z - jI_i Z_0 \sin \beta z$	
Current along line $I(z)$	$I_i \cosh \gamma z - \frac{V_i}{Z_0} \sinh \gamma z$	$I_i \cos \beta z - j\frac{V_i}{Z_0} \sin \beta z$	
Reflection coefficient K_R	$\frac{Z_L - Z_0}{Z_L + Z_0}$	$\frac{Z_L - Z_0}{Z_L + Z_0}$	
Standing wave ratio	$\frac{1 + \lvert K_R \rvert}{1 - \lvert K_R \rvert}$	$\frac{1 + \lvert K_R \rvert}{1 - \lvert K_R \rvert}$	

R, L, G, C Distributed resistance, inductance, conductance, capacitance per unit length.
l Length of line.
Subscript i denotes input end quantities.
Subscript L denotes load end quantities.

z Distance along line from input end.
λ Wavelength measured along line.
v Phase velocity of line equals velocity of light in dielectric of line for an ideal line.

2

THE EQUATIONS OF STATIONARY ELECTRIC AND MAGNETIC FIELDS

2.01 Introduction

Static electric and magnetic fields are of great interest to radio engineers. The equations describing these fields are probably used more often than those of varying fields. One reason for this is that even at very high frequencies, the electromagnetics of static fields suffices to explain many phenomena and yield sufficiently accurate quantitative results. In still other high-frequency problems, the distribution of the fields may be exactly the same as in certain other static problems. By reviewing these familiar static fields, we also hope to understand better the philosophy of all fields and circuits and to introduce tools which will prove invaluable when we are confronted with changing field problems.

The first few chapters of this text will present a set of equations which will include all the knowledge of electromagnetics necessary to solve most radio engineering electromagnetic problems: radiation from an antenna, propagation of waves in space and along transmission lines, the special case of static fields, etc. Now, we might quite logically present merely a set of general differential equations at the beginning as the fundamentals of electricity and magnetism. From these could then be derived the relations applicable to fields which do not vary with time, and the results could be checked with knowledge and ideas of static fields. Such an approach will actually be used later in obtaining the low-frequency circuit equations (Kirchhoff's laws) from the general differential equations, but the opposite approach will be more valuable here; that is, we shall take familiar experimental results in electrostatics as the starting point and seek to derive the most useful equations that will describe the experimental results.

If we begin the study of fields from some observed law which may be regarded as fundamental, the statement of the law should be made as general as possible so that it will be useful to describe a variety of conditions. We should always be critical of this procedure, since these " laws " represent generalizations from several experiments, all of which are special in nature; there is nothing to assure us, in extending them from the range of magnitudes and conditions in which they were determined to an entirely new set of magnitudes and conditions, that the

phenomena predicted will actually ever be observed until these too are checked by special experiment. Once checked, the derived form of the relation might just as well have been the fundamental law. In fact we might have started from that point had we thought of it first. In applied physics it is particularly necessary to find a large number of these derived relations, since it is seldom convenient to use the law in its original form for all design or analysis. There are often so many of these forms that the engineer in using a dozen different relations for as many separate problems may be quite unaware that many of these relations are in reality equivalent.

To study static fields some experimental " laws " will be taken as fundamental. By transformations, definitions, and generalizations, other forms of the law will be obtained, which may be more general or more convenient to use for certain problems. We shall extend the laws developed from macroscopic systems to the infinitesimal, and so obtain differential equations with which we may study continuous variations from point to point, as well as discrete systems. Once this extension has been justified, the differential equation will be the most valuable tool for the study of fields.

As the discussion proceeds, it will be noticed that directions appear as frequently as magnitudes in the statement of the laws, so that quite naturally it will be necessary to use a short-cut vector notation to save time, space, and many words. It will soon be discovered that this notation permits many short cuts in manipulation, and, most important of all, leads to a very superior way of thinking about electric and magnetic effects.

STATIC ELECTRIC FIELDS

2.02 The Problem of Static Electric Fields

The problem which must be solved in static electric field theory is that of obtaining relations which involve the geometrical configurations of conductors and dielectrics, the distribution of charges on the conductors and in the dielectric medium separating them, the potential differences between conductors, and the field distribution in the dielectric. Several or all of these factors will enter into the determination of capacitance between conductors, the maximum gradient in insulation, the amount of field between deflecting plates in an oscilloscope, the amount of shielding which a grid provides in a vacuum tube, or the accelerating force on an electron in an electron gun.

Essentially, the problem is one of equilibrium. We require a knowledge of the forces that act on charges, thus making them move to even-

tual equilibrium positions, and we must know the manner in which conductors and dielectrics affect the charge distribution and the field distribution.

2.03 Force between Electric Charges; Electrostatic Units

We shall take as the starting point for electrostatics the experimental law of Coulomb, which gives the force between two electric charges. The law includes the following information:

1. Like charges repel, opposites attract.
2. Force is proportional to the product of charge magnitudes.
3. Force is inversely proportional to the square of the distances between charges.
4. Force is dependent upon the medium in which the charges are placed.
5. Force acts along the line joining the charges.

This information may be written as an equation.

$$f = -k\frac{q_1q_2}{\epsilon r^2} \qquad [1]$$

In this equation, f is defined as the force of attraction acting on the line between charges, q_1 and q_2 represent the charges in magnitude and sign, r is the distance between charges, ϵ is a property of the medium which may be called the dielectric constant, and k is a constant of proportionality which must be included for the present, since we have not as yet defined units.

The equation may be written so that the direction of the force is included.

$$\bar{f} = -k\frac{q_1q_2}{\epsilon r^2}\bar{a}_r \qquad [2]$$

The bar above f denotes that force is a directed quantity, or vector. That is, it has both magnitude and direction. The magnitude is given by the numerical value of $k\frac{q_1q_2}{\epsilon r^2}$ which in itself implies no direction, and it is accordingly called a scalar quantity. The direction of $\bar{f}$ is given by $\bar{a}_r$, a vector of unit length pointing from one charge directly toward the other, and the sign of $-q_1q_2$. Thus if q_1 and q_2 have opposite signs, $-q_1q_2$ is positive, $\bar{f}$ has exactly the direction of $\bar{a}_r$, and the force is from one charge directly toward the other. If q_1 and q_2 have the same sign, $-q_1q_2$ is negative, $\bar{f}$ has exactly the opposite direction to $\bar{a}_r$, and the force is from one charge directly away from the other. This is merely the statement of opposite charges attracting, likes repelling.

Vectors such as $\bar{a}_r$ are known as unit vectors and will be useful throughout the study of fields, since they serve to indicate direction without interfering with magnitudes.

Equation (1) may be made to define a system of units. If a unit charge is defined as that charge which repels an exactly similar charge with a force of 1 dyne when the two are placed 1 cm apart in a vacuum, a system of units is defined such that $k/\epsilon = 1$. In this system of units, the dielectric constant of vacuum is further defined as unity, so that k is also unity. For reasons which appear when other units are considered, dielectric constants referred to vacuum as unity will be denoted by ϵ'.

The system based upon the above definitions is known as the electrostatic system of units (esu), and the unit of charge is called the statcoulomb. It will be the basis for all subsequent definitions and equations encountered in the study of electrostatics in this chapter. Later, in the study of magnetic fields, it will be convenient to define a new system. In still later problems a third system, a practical system of units, will be used, which will actually be our preferred system. It is necessary to understand all these systems well if the engineer is to use the present reference books on electricity and magnetism with ease, for formulas are given in the system most convenient, or at least what appears most convenient to that author. To keep confusion to a minimum, the formulas of the text will eventually be restricted to the practical system after the various systems have been explained and discussions reduce to the matter of solving practical problems.

2.04 Electric Field Intensity

If the unit charge as now defined is placed at a distance r from a charge q in vacuum, the force law shows that it experiences a force q/r^2 dynes. In the more general case, any charge placed in the vicinity of a system of charges experiences a force whose magnitude and direction are functions of the amounts and positions of all charges of the system. A region so influenced by charges is called a region of electric field. The force per unit charge on a positive test charge at a point is defined as the strength of electric field or electric intensity at the point, provided the test charge is so small that it does not disturb the original charge distribution of the system. Since the force on the test charge has direction as well as magnitude, the electric intensity is a vector. The electric intensity or electric field vector is then defined by

$$\bar{E} = \frac{\bar{f}}{\Delta q} \qquad [1]$$

where $\bar{f}$ is the force acting upon the infinitesimal test charge, Δq.

The electric field intensity arising from a point charge, q, in any medium is given by the force law Eq. 2.03(2). In esu,

$$\bar{E} = \frac{q}{\epsilon' r^2}(-\bar{a}_r) \tag{2}$$

Since $\bar{a}_r$ is the unit vector directed from the point toward the charge q, $-\bar{a}_r$ is directed away from the charge, so that the electric field vector points away from positive charges and toward negative charges.

2.05 Displacement Flux

Equation 2.04(2) shows that the electric intensity is dependent upon the medium in which the charge is placed. Suppose a new vector quantity is defined, independent of the medium. Define the displacement $\bar{D}$ by

$$\bar{D} = \epsilon' \bar{E} \tag{1}$$

The displacement for the point charge then becomes

$$\bar{D} = \frac{q}{r^2}(-\bar{a}_r) \tag{2}$$

The displacement at any point is thus a function of charge and position alone; consequently the charge may be thought of as giving rise to so much displacement in its surrounding medium. Each charge may be considered as a source of displacement flux or lines of flow in the medium, and according to this concept, $\bar{D}$ is an electric flux density, with the important property that, unlike electric intensity, it is independent of medium.

Take, for example, an imaginary spherical surface with charge $+q$ at its center. At each point on the sphere there are q/r^2 lines of displacement flux per unit area passing radially outward, so that emanating from the entire sphere there are $4\pi q$ lines, regardless of the size of sphere or the medium in which the charge is placed.

2.06 Gauss's Law

In Art. 2.05, it was found that the total flux emanating from a sphere of any radius with charge q at the center was $4\pi q$ lines. As a first step in reducing Coulomb's law and the accompanying definitions to most useful form, it will be helpful to become more general. Consider a volume of any shape containing charges. If one of these point charges, q, is considered (Fig. 2.06), the field intensity and displacement can be calculated for any point on the surface by equations of previous articles. Thus at point P, D is q/r^2. (When a quantity normally a vector

appears without the bar, it signifies that magnitude alone is being considered.) If θ is the angle between $\bar{D}$ and the normal to the surface at P, the amount of flux passing through an elemental surface dS is

$$d\psi = \frac{q}{r^2}\, dS \cos\theta$$

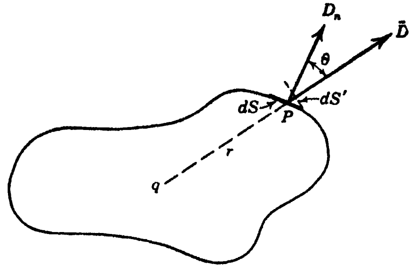


FIG. 2.06. Charge q and arbitrary surrounding surface.

$dS \cos\theta$ is the area dS', the component of dS normal to $\bar{D}$. From the definition of solid angle, the solid angle $d\Omega$ subtended by either dS or dS' is dS'/r^2; so then the amount of flux passing through the elementary surface is $q\,d\Omega$. To obtain total displacement flux, this expression is integrated over all the surface, which amounts to integrating $d\Omega$. The result is then $4\pi q$. So

$$\oint_S D \cos\theta\, dS = 4\pi q \qquad [1]$$

D is the magnitude of the displacement at any point on the surface, dS is an elemental area at that point, and θ is the angle between the displacement vector and the normal to the surface. $\int_S$ is used to denote a surface integral, the integral of a quantity over a given surface.

If the elemental surface dS is represented by a vector, the equation may be written more simply. Define $d\bar{S}$ as the vector which represents that elemental area. Its magnitude is dS, and its direction is that of the outward normal to the surface. Then replace $D\, dS \cos\theta$ by $\bar{D} \cdot d\bar{S}$.

The product just defined is called the dot product of two vectors, or the scalar product, since it results by definition in a scalar $D\, dS \cos\theta$, which is the product of the magnitude of one vector by the projection of the other upon it. Gauss's law of (1) may then be written

$$\oint_S \bar{D} \cdot d\bar{S} = 4\pi q \qquad [2]$$

The q considered was only one of the charges of the system, but since it might have been located at any point inside the surface, and since fields arising from several charges may be superposed, the q of (2) may be considered as the sum of all charges enclosed by the system. In particular, when the charge is distributed throughout the region with a charge density, or charge per unit volume, ρ, at any point, the total charge

enclosed is the volume integral of this charge density. Equation (2) is then written

$$\oint_S \bar{D} \cdot d\bar{S} = 4\pi \int_V \rho \, dV \qquad [3]$$

$\int_V$ denotes volume integral, or integral of a quantity throughout a given volume.

Rigorous proofs for Gauss's law may be found in several of the references,[1] but this demonstration should give a clear picture of the concept of flux. Once one has learned to think of lines of flux emanating from a point charge radially in all directions, then it seems that the amount of flux passing through any imaginary enclosing surface must be constant, since no flux lines are created or destroyed in passage.

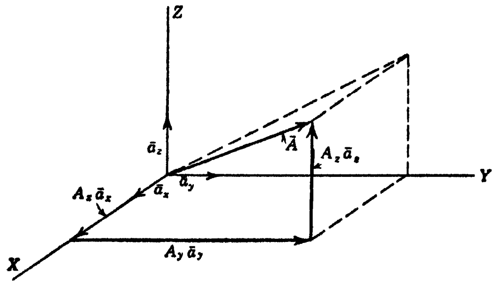


FIG. 2.07. Vector $\bar{A}$ and its rectangular components.

2.07 Scalar or Dot Product of Vectors

The vector operation defined in the last article is important since there is often occasion to multiply one vector by the projection of the other upon it. That is, if $\bar{A}$ and $\bar{B}$ are vectors (of magnitudes A and B) with an angle of θ between them, $AB \cos \theta$ is of interest. This has been written as $\bar{A} \cdot \bar{B}$. (Read A dot B.) This product may now be expressed in terms of the components of $\bar{A}$ and $\bar{B}$ along the coordinate axes.

A unit vector has already been defined in the statement of Eq. 2.03(2). If $\bar{a}_x$, $\bar{a}_y$, $\bar{a}_z$ are three such unit vectors having the directions of the three axes in rectangular coordinates, and if A_x, A_y, and A_z are the magnitudes of the components of $\bar{A}$ along these axes, $\bar{A}$ may be written

$$\bar{A} = A_x\bar{a}_x + A_y\bar{a}_y + A_z\bar{a}_z$$

The addition of the three component vectors to obtain $\bar{A}$ is performed according to ordinary engineering ideas of vector addition (Fig. 2.07).

[1] In this text we shall mean the general references of Appendix A, if specific references are not given in a footnote.

The dot product is

$$\bar{A} \cdot \bar{B} = (A_x\bar{a}_x + A_y\bar{a}_y + A_z\bar{a}_z) \cdot (B_x\bar{a}_x + B_y\bar{a}_y + B_z\bar{a}_z)$$

If the multiplication is carried through term by term, with the dot product between component vectors retained:

$$\bar{A} \cdot \bar{B} = A_xB_x\bar{a}_x \cdot \bar{a}_x + A_xB_y\bar{a}_x \cdot \bar{a}_y, \text{ etc.}$$

The terms $\bar{a}_x \cdot \bar{a}_x$, $\bar{a}_y \cdot \bar{a}_y$, $\bar{a}_z \cdot \bar{a}_z$ are unity by definition of the unit vectors and the dot product. The terms $\bar{a}_x \cdot \bar{a}_y$, $\bar{a}_y \cdot \bar{a}_z$, etc., are zero since the angle between any of these unit vectors and either of the other two is 90°. The scalar product then reduces to

$$\bar{A} \cdot \bar{B} = A_xB_x + A_yB_y + A_zB_z \qquad [1]$$

2.08 Tubes of Flux

The concept of flux passing through an area obviously does not have to be limited to electric phenomena. If $\bar{D}$ is any vector function of space, the product of the magnitude of $\bar{D}$ at any point by an element of area perpendicular to $\bar{D}$ at that point may be called the flux of $\bar{D}$ passing through that area. The total flux flowing through a surface is given by the surface integral

$$\psi = \int_S \bar{D} \cdot d\bar{S} \qquad [1]$$

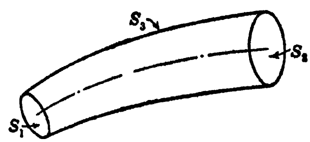


FIG. 2.08. Tube of flux.

As before, the convention is to regard the vector representing the elements of area as pointing outwards.

Consider a surface (Fig. 2.08), bounded by two planes, S_1 and S_2, perpendicular to the field vector at two points, and a surface S_3 always parallel to the direction of the field vector. If there is no charge enclosed, Gauss's law gives

$$\int_{S_1} \bar{D} \cdot d\bar{S} + \int_{S_2} \bar{D} \cdot d\bar{S} + \int_{S_3} \bar{D} \cdot d\bar{S} = 0 \qquad [2]$$

Since S_3 is always parallel to $\bar{D}$, there is no flux flowing out through S_3.

$$\int_{S_3} \bar{D} \cdot d\bar{S} = 0 \qquad [3]$$

So

$$\int_{S_1} \bar{D} \cdot d\bar{S} = -\int_{S_2} \bar{D} \cdot d\bar{S} \qquad [4]$$

This equation states that the flux passing through the plane S_1 is that which comes out of the plane S_2, so that total flux across any cross section of the tube is a constant. Such a tubular region may be called a tube of flux. To study the field intensity distribution, it is sometimes helpful to draw out many of these tubes, the size of area being so selected that the flux through the area is one unit. Lines are often used to represent the tubes, and the tubes loosely called lines. Thus the smaller the spacing of these lines, i.e., the more lines per square centimeter, the stronger the flux density at that point.

We have agreed that lines of flux emanate from charges and are continuous in regions without charge; consequently the lines of force for an electrostatic field must begin on a charge and end on a charge. To be consistent with the convention already adopted, the line is said to begin on a positive charge and end on a negative charge.

2.09 Charges on Conductors

If expressions for the field in differential equation form are to be obtained, it is important that boundary conditions for application to their solutions be well understood. Conducting metal surfaces will often form these boundaries.

Conductors are defined as those materials which readily permit a current flow, or motion of charges. So if charges are placed on or in conductors, they will move about as long as there is the slightest electric field producing a force upon them. After they have reached equilibrium, the necessary condition for a static field to exist, all the electric field inside the conductor or tangential to its surface must have disappeared. If there were charges in the body, Gauss's law would require an electric field in the vicinity of these charges, so that this is an impossible condition for the static case. All the charge in electrostatics must then reside on the surface and must be distributed so that the component of electric field intensity tangential to the surface and the total electric field intensity inside the material surface of the conductor are zero.

2.10 Boundary between Conductors and Dielectrics

Determination of the charge on a conductor which bounds a given electric field demonstrates the application of Gauss's law to a case much more difficult to study directly from Coulomb's law. Consider the imaginary surface indicated in cross section by the dotted lines of Fig. 2.10. There can be no flux through the surface dS' since it is submerged in the metal, below the surface. The distance h can be made as small as we like compared with dS, since dS must be only an infinitesimal distance outside the metal, dS' must be only an infinitesimal dis-

tance inside. So then the flux passing through all surfaces containing h can be made negligible. This leaves the total flux to pass through the surface dS. By Gauss's law this flux must be 4π times the charge enclosed, which is $\rho_s\, dS$. (ρ_s is the surface charge density, or charge per unit area.) Thus

$$\epsilon' E\, dS = 4\pi\rho_s\, dS$$

or

$$\epsilon' E = 4\pi\rho_s \qquad [1]$$

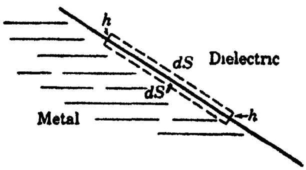


FIG. 2.10. Cross section, showing surface separating a dielectric and a conductor.

E is the electric intensity normal to the surface, which is the total E in the present case since there can be no component tangential to the surface of the conductor. The result is directly due to the requirement that 4π lines of flux leave each unit positive charge, and that $\bar{D}$ and $\bar{E}$ are normal to the conductor surface external to the conductor, zero inside the conductor. The result may be used to find the amount of electric flux leaving a conductor at every point if the charge distribution on the conductor is known, or conversely, to evaluate the charge that must be induced on a conductor at every point when a known distribution of electric field ends on this conductor.

2.11 Diverging and Converging of Flux Lines

Gauss's law was derived from Coulomb's law which was determined by experiment on systems of finite size. Let us extend it to an infinitesimally small system. Equation 2.06(3) may then be written:

$$\lim_{\Delta V \to 0} \frac{\oint_S \bar{D} \cdot d\bar{S}}{\Delta V} = \lim_{\Delta V \to 0} \frac{4\pi \oint_V \rho\, dV}{\Delta V} \qquad [1]$$

The right side is, by inspection, merely $4\pi\rho$. The left side is the amount of displacement flux per unit volume flowing out of an infinitesimal volume. This will be defined as the divergence of displacement, abbreviated div $\bar{D}$. Then

$$\text{div}\, \bar{D} = 4\pi\rho \qquad [2]$$

To make the picture clearer, consider the infinitesimal volume as a rectangular parallelepiped of dimensions Δx, Δy, Δz as shown in Fig. 2.11. To compute the amount of flux leaving such a volume element as compared with that entering it, note that the flux passing through any face of the parallelepiped can differ from that which passes through the

opposite face only if the displacement perpendicular to those faces varies from one face to the other. If the distance between the two faces is small, then to a first approximation the difference in any vector function on the two faces will simply be the rate of change of the function with distance times the distance between faces.

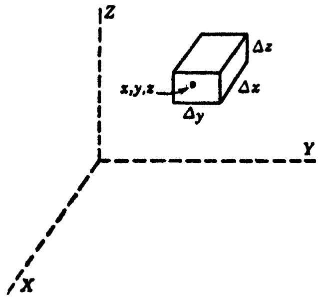


FIG. 2.11.

If the displacement vector at the center has components D_x, D_y, D_z,

$$D_x\Big|_{x+\frac{\Delta x}{2}} = D_x + \frac{\Delta x}{2}\frac{\partial D_x}{\partial x}$$

$$D_x\Big|_{x-\frac{\Delta x}{2}} = D_x - \frac{\Delta x}{2}\frac{\partial D_x}{\partial x} \tag{3}$$

The displacement flux flowing out the front face is $\Delta y\, \Delta z\, D_x\Big|_{x+\frac{\Delta x}{2}}$, and that flowing in the back face is $\Delta y\, \Delta z\, D_x\Big|_{x-\frac{\Delta x}{2}}$, leaving a net flow out of $\Delta x\, \Delta y\, \Delta z\, \dfrac{\partial D_x}{\partial x}$. Similarly for the y and z directions, so that net flux flow out of all the parallelepiped is

$$\Delta x\, \Delta y\, \Delta z\, \frac{\partial D_x}{\partial x} + \Delta x\, \Delta y\, \Delta z\, \frac{\partial D_y}{\partial y} + \Delta x\, \Delta y\, \Delta z\, \frac{\partial D_z}{\partial z}$$

By Gauss's law, this must be $4\pi\rho\, \Delta x\, \Delta y\, \Delta z$. So, in the limit,

$$\frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} + \frac{\partial D_z}{\partial z} = 4\pi\rho \tag{4}$$

An expression for div $\bar{D}$ in rectangular coordinates is obtained by comparing (2) and (4).

$$\operatorname{div} \bar{D} = \frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} + \frac{\partial D_z}{\partial z} \tag{5}$$

It will be convenient to define a vector operator ∇ (pronounced del) in rectangular coordinates as

$$\nabla = \bar{a}_x \frac{\partial}{\partial x} + \bar{a}_y \frac{\partial}{\partial y} + \bar{a}_z \frac{\partial}{\partial z} \tag{6}$$

Consider the expansion for the dot or scalar product, Eq. 2.07(1), and the definition of ∇ above. Then (5) indicates that div $\bar{D}$ can correctly be written as $\nabla \cdot \bar{D}$. It should be remembered that ∇ is not a true

vector but rather a vector operator. We need not worry about its meaning except when it is operating on another quantity in a defined manner. The divergence represents the first of several of these operations to be defined.

$$\nabla \cdot \bar{D} = \frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} + \frac{\partial D_z}{\partial z} \tag{7}$$

Finally

$$\nabla \cdot \bar{D} \equiv \text{div } \bar{D} = 4\pi\rho \tag{8}$$

The divergence is made up of space derivatives of the field, so (8) is evidently a differential equation derived by generalizing from the previous laws for comparatively large systems. It will be so important that we should become accustomed to looking at it as an expression for Gauss's law generalized to a point in space. The physical significance of the divergence must be clear. It is, as defined, a description of the manner in which a field varies at a point. It is the amount of flux per unit volume emerging from an infinitesimal volume at a point. With this picture in mind, (8) seems a logical extension of Gauss's law. Since Gauss's law was in turn derived from Coulomb's force law, the above equation may be considered as a differential equivalent of Coulomb's law.

2.12 Divergence Theorem

If Eq. 2.11(2) is integrated over any volume,

$$\int_V \text{div } \bar{D}\, dV = 4\pi \int_V \rho\, dV \tag{1}$$

Replace the last term by its equivalent from Gauss's law, Eq. 2.06(3).

$$\int_V \text{div } \bar{D}\, dV = \int_S \bar{D} \cdot d\bar{S} \tag{2}$$

Although this relation has been derived from a consideration of $\bar{D}$, a little thought will show that it is a direct consequence of the definition of divergence and so must hold for any vector field. For if divergence of any vector is considered as a density of outward flux flow from a point for that vector, then it seems that the total outward flux flow from a closed region must be obtained by integrating the divergence throughout the volume. If $\bar{F}$ is any vector

$$\int_V \text{div } \bar{F}\, dV = \int_V \nabla \cdot \bar{F}\, dV = \int_S \bar{F} \cdot d\bar{S} \tag{3}$$

This relation is known as the divergence theorem or Gauss's theorem (as

distinguished from Gauss's law of Art. 2.06) and will be useful later in manipulating vector equations in order to arrive at their most useful forms. Note that the theorem is true for any continuous vector function of space, regardless of the physical significance of that vector.

2.13 Conservative Property of Electric Fields

Before proceeding very far in attempts to build up pictures and quantitative relations for electrostatic fields, we should pause to look into the very important matter of energy. The field may be checked with ideas of conservation of energy, to determine, for example, whether the energy of the electrostatic field is a function merely of its state at any given time, or whether it depends upon the manner in which that state occurred. We no doubt already feel certain that the energy of an electrostatic field depends only upon the amounts and positions of the charges, and not on how they grew; the inverse square law tells us that this must be so.

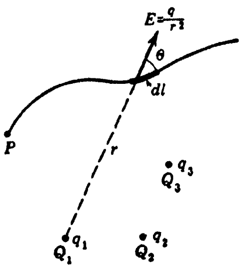


FIG. 2.13.

The force on a small charge Δq moved from infinity to a point P in the vicinity of a system of charges: q_1 at Q_1, q_2 at Q_2, q_3 at Q_3, etc., may be calculated at any point along its path. Consider, for example, the work integral arising from q_1. The work is the integral of force component in the direction of the path, multiplied by differential path length (Fig. 2.13).

$$U_1 = -\int \frac{\Delta q q_1 \cos\theta\, dl}{\epsilon' r^2}$$

But $dl \cos\theta$ is dr, so the integral is simply

$$U_1 = -\int_\infty^{PQ_1} \frac{\Delta q q_1\, dr}{\epsilon' r^2}$$

Similarly, for contributions from other charges, so that the total work integral is

$$U = -\int_\infty^{PQ_1} \frac{\Delta q q_1}{\epsilon' r^2}\, dr - \int_\infty^{PQ_2} \frac{\Delta q q_2}{\epsilon' r^2}\, dr - \int_\infty^{PQ_3} \frac{\Delta q q_3}{\epsilon' r^2}\, dr \cdots$$

Integrating,

$$U = \frac{(\Delta q) q_1}{\epsilon' PQ_1} + \frac{(\Delta q) q_2}{\epsilon' PQ_2} + \frac{(\Delta q) q_3}{\epsilon' PQ_3} + \cdots \qquad [1]$$

Equation (1) shows that the work done is only a function of final positions and not of the path of the charge. This conclusion leads to

another: if a charge is taken around any closed path, no net work is done. Mathematically this is written

$$\oint \bar{E} \cdot \overline{dl} = 0 \qquad [2]$$

In the study of magnetic fields, we shall find corresponding work integrals which are not zero.

2.14 Electrostatic Potential

To solve the differential field equations, it is often convenient to introduce mathematical tools known as potential functions, which may aid materially during the solution but which need not appear in the final result. It is never necessary to give these mathematical tools physical significance though it often may be desirable. We are already quite familiar with the potential function of electrostatics, and in this case it may easily have more significance for us than the fields, which were themselves only defined concepts to describe the situation in a region containing charges.

The common potential function in electrostatics is a scalar quantity defined so that the difference in this function between two points P and Q is given by the integral

$$\Phi_P - \Phi_Q = -\int_Q^P \bar{E} \cdot \overline{dl} \qquad [1]$$

The physical significance that may be attached to it is now apparent, for (1) is an expression for the work done on a unit charge in moving it from P to Q. The conclusion of the preceding article that the work in moving around any closed path is zero shows that the potential function defined is single valued; that is, corresponding to each point of the field there is only one value of potential, though the potential may, of course, vary from point to point.

Only a *difference* of potential has been defined. The potential of any point can be arbitrarily fixed and then the potentials of all other points in the field found by application of the definition to give potential differences between all points and the base. This base is quite arbitrary since the potential differences alone have significance. For example, in certain cases it may be convenient to define the potential at infinity as zero and then find the corresponding potentials of all points in the field; for the determination of the field between two conductors, it will be more convenient to select the potential of one of these as zero.

If the potential at infinity is taken as zero, it is evident that the potential at the point P in the system of charges, Art. 2.13, is given by

U of Eq. 2.13(1) divided by Δq so

$$\Phi = \frac{q_1}{\epsilon' PQ_1} + \frac{q_2}{\epsilon' PQ_2} + \frac{q_3}{\epsilon' PQ_3} + \cdots \tag{2}$$

Generalizing to the case of continuously varying charge density,

$$\Phi = \int_V \frac{\rho\, dV}{\epsilon' r} \tag{3}$$

ρ is the charge density, and the integral signifies that a summation should be made similar to that of (2) but continuous over all space. There are, of course, arbitrary added constants if the potential at infinity is not taken as zero.

At once there is evidence of the usefulness of the potential tool, for Φ is obtained by simple scalar addition; it would have been necessary to perform corresponding vector additions to obtain fields directly. Since the fields can be obtained simply from the potential, the work of obtaining electric fields from charges is simplified. We shall next show how this may be done.

2.15 Gradient

If the definition of potential difference is applied to two points a distance dl apart,

$$d\Phi = -\bar{E} \cdot \overline{dl} \tag{1}$$

$\overline{dl}$ may be written in terms of its components and the defined unit vectors (Art. 2.07).

$$\overline{dl} = dx\, \bar{a}_x + dy\, \bar{a}_y + dz\, \bar{a}_z \tag{2}$$

Expand the dot product according to Eq. 2.07(1)

$$d\Phi = -(E_x\, dx + E_y\, dy + E_z\, dz)$$

Since Φ is a function of x, y, and z, the total derivative may also be written

$$d\Phi = \frac{\partial \Phi}{\partial x} dx + \frac{\partial \Phi}{\partial y} dy + \frac{\partial \Phi}{\partial z} dz$$

From a comparison of the two expressions,

$$E_x = -\frac{\partial \Phi}{\partial x}$$

$$E_y = -\frac{\partial \Phi}{\partial y} \tag{3}$$

$$E_z = -\frac{\partial \Phi}{\partial z}$$

$$\bar{E} = -\left(\bar{a}_x \frac{\partial \Phi}{\partial x} + \bar{a}_y \frac{\partial \Phi}{\partial y} + \bar{a}_z \frac{\partial \Phi}{\partial z}\right) \quad [4]$$

or

$$\bar{E} = -\text{grad } \Phi \quad [5]$$

where grad Φ, an abbreviation of the gradient of Φ, is a vector showing the direction and magnitude of the maximum space variation in the scalar function Φ, at any point in space. It is the maximum variation that is represented because the gradient is the vector sum of the variations in all three directions.

The vector operator ∇ was defined by Eq. 2.11(6). Then grad Φ may be written as $\nabla\Phi$ if the operation is interpreted

$$\nabla\Phi = \bar{a}_x \frac{\partial \Phi}{\partial x} + \bar{a}_y \frac{\partial \Phi}{\partial y} + \bar{a}_z \frac{\partial \Phi}{\partial z} \quad [6]$$

and

$$\bar{E} = -\text{grad } \Phi \equiv -\nabla\Phi \quad [7]$$

Problem 2.15. Demonstrate that the gradient of Φ, as defined by Eq. 2.15(6), is indeed a vector representing magnitude and direction of the maximum space variation of Φ.

2.16 Equipotentials

All points of a field having the same potential may be thought of as connected by equipotential surfaces. The distribution and spacing of these equipotential surfaces can be used to describe the field. The electric field vector must be perpendicular to these surfaces at every point, for if there were the slightest component tangential to the surface, say E_t, then two points $d\xi$ apart would have a potential difference $E_t d\xi$ which would violate the condition for an equipotential surface. This was the same requirement considered in Art. 2.09 for conducting surfaces, so it follows at once that all conducting surfaces in electrostatics must be equipotential surfaces.

If the potential were to vary in one direction only, say x, as in a potential difference applied between two infinite parallel conducting planes perpendicular to the x axis, the electric field, or negative gradient of potential, would be entirely in the x direction. The equipotential surfaces would be perpendicular to the x axis, or parallel to the conducting planes, as would be expected from symmetry.

In the general case the field will vary in all directions, the equipotential surfaces will not be planes, and the gradient will be made up of components in the x, y, and z directions having magnitudes proportional to the variation of potential in those three directions.

2.17 Laplace's and Poisson's Equations

It will often be convenient to work directly with potentials instead of fields, since the specified conditions of the problem, i.e., the boundary conditions, may be given in terms of potentials (say the voltage applied between two conductors).

If the value of $\bar{E}$ from Eq. 2.15(7) is substituted in Eq. 2.11(2), and if ϵ' is constant throughout the region,

$$-\text{div (grad } \Phi) = -\nabla \cdot \nabla\Phi = \frac{4\pi\rho}{\epsilon'}$$

But from the equations for divergence and gradient in rectangular coordinates [Eqs. 2.11(7) and 2.15(6)]

$$\nabla \cdot \nabla\Phi = \frac{\partial^2\Phi}{\partial x^2} + \frac{\partial^2\Phi}{\partial y^2} + \frac{\partial^2\Phi}{\partial z^2} \qquad [1]$$

So

$$\frac{\partial^2\Phi}{\partial x^2} + \frac{\partial^2\Phi}{\partial y^2} + \frac{\partial^2\Phi}{\partial z^2} = -\frac{4\pi\rho}{\epsilon'} \qquad [2]$$

This is the differential equation which relates potential variation at any point to the charge density at that point. It is known as Poisson's equation and is often written

$$\nabla^2\Phi = -\frac{4\pi\rho}{\epsilon'} \qquad [3]$$

where

$$\nabla^2\Phi \equiv \nabla \cdot (\nabla\Phi) \equiv \text{div (grad } \Phi) \qquad [4]$$

In the special case of a charge-free region, Poisson's equation reduces to

$$\frac{\partial^2\Phi}{\partial x^2} + \frac{\partial^2\Phi}{\partial y^2} + \frac{\partial^2\Phi}{\partial z^2} = 0$$

or

$$\nabla^2\Phi = 0 \qquad [5]$$

which is known as Laplace's equation.

Any number of possible configurations of potential surfaces will satisfy the requirements of (3) and (5). All are called solutions to these equations. It is necessary to know the conditions existing around the boundary of the region in order to select the particular solution which applies to a given problem. It can be shown mathematically that once ρ is given at every point in a region and Φ is given at every point on the surface surrounding the region, only one potential distribution is possible.

Equations exactly similar in form to (3) and (5) are found in many branches of physics. In fact we shall discover later that they are true not only when the function is a static potential; for example, the function may be the static field strength vectors or certain of their components. Laplace's and Poisson's equations are of first importance in getting answers to all problems in which static electric and magnetic effects are involved. The ability to choose solutions of these equations is fundamental in arriving at the final solutions to the common problems discussed in Art. 2.02. For that reason the next chapter will be devoted almost entirely to a discussion of the techniques of building up solutions to these equations to fit boundary conditions that are likely to occur in practical problems.

STATIC MAGNETIC FIELDS

2.18 Magnetic Field of a Direct Current

In the first part of this chapter the concept of the electric field was developed from the experimental observation that a charge in the vicinity of other charges experiences a force. Experimentally, it can also be determined that a loop carrying current will be acted on by a force if it is brought in the vicinity of another current or system of currents. The region in which such forces exist is spoken of as a region of magnetic field. Now, of course, the study of magnetism may be approached from various standpoints; we shall find it most advantageous to study it as an effect due to current flow. In this chapter we shall limit ourselves to a discussion of the concepts of magnetic fields due to unchanging currents, i.e., static currents, just as earlier in the chapter the discussion was limited to the electric field effects of static charges. As before, the conclusions and concepts will be applicable not only to static currents, but also to low-frequency problems and to many high-frequency situations in which the field distribution will later be shown to be identical with those of statics.

Experimental measurements show that the force between two or more static current elements is dependent upon the following factors.

1. Direction of current flow.
2. Magnitude of currents.
3. Distances between currents.
4. Orientation of currents.
5. Shape and size of current path.

A comparison of these and the corresponding factors for electric charges in Art. 2.03 warns us that the force law for currents will not be written so simply as that for charges. Current is a vector, i.e., it has direction, to mention one complicating factor; the orientation of the current vectors must appear in the force law. It will be convenient therefore to define a field strength for magnetic fields before attempting to write the force law. The electric field intensity was defined in terms of force on a small charge. The magnetic intensity, or magnetic field vector $\bar{H}$, is defined in terms of the force on a small current element such that

$$df = I\,dl\mu' H \sin\theta \qquad [1]$$

μ' is a function of the medium known as permeability, I is the current flowing in the element dl, H is the magnitude of the magnetic intensity, θ the angle between $\overline{dl}$ and $\bar{H}$, and df the magnitude of the force on the current element. The direction of $\overline{df}$ is along the perpendicular to the plane containing $\overline{dl}$ and $\bar{H}$ and in the direction of advance of a right-hand screw if $\overline{dl}$ is rotated into $\bar{H}$. This equation enables one to measure field strength at any point, this field strength presumably arising from a distribution of currents in the neighborhood, although, of course, it may be due to permanent magnets.

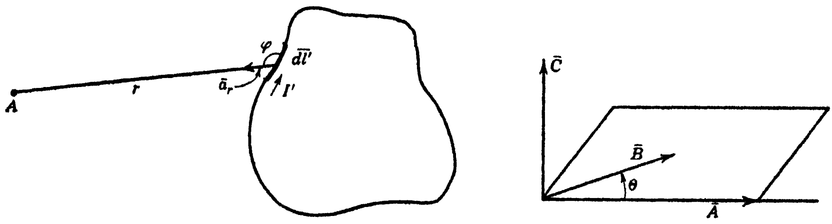
The remainder of the information obtained experimentally is contained in a second law relating the field to the currents which produce it. Although this law is probably correctly credited to Biot, it is more commonly known as Ampère's law, so we shall use that designation. The law is

$$dH = \frac{I'\,dl' \sin\phi}{r^2} \qquad [2]$$

As in Fig. 2.18*a*, dl' is a contributing current element having current I', r is the magnitude of the vector from the element to the point at which H is to be determined, ϕ is the angle between dl' and r, and dH is the magnitude of the contribution to H from the element dl'. The direction of $\overline{dH}$ is given by the normal to the plane containing $\overline{dl}'$ and $\bar{r}$, and by the direction of advance of a right-hand screw if $\overline{dl}'$ is rotated into $\bar{r}$.

If a simple vector notation is introduced, these laws may be written clearly in a vector form that includes these clumsy direction laws. Both

laws involve a vector which is perpendicular to the plane containing two other vectors and has as magnitude the product of the magnitude of one and the component of the second perpendicular to the first. Thus if $\bar{C}$ is the vector resulting from the combination of two such vectors $\bar{A}$



(a) (b)

FIG. 2.18.

and $\bar{B}$ (Fig. 2.18*b*), define $\bar{C} = \bar{A} \times \bar{B}$ as a vector product of the two vectors $\bar{A}$ and $\bar{B}$. $\bar{C}$ is a vector perpendicular to the plane containing $\bar{A}$ and $\bar{B}$, of magnitude $AB \sin \theta$ and having a direction given by the direction of advance of a right-hand screw if $\bar{A}$ is rotated into $\bar{B}$. From this definition it is seen that

$$\bar{A} \times \bar{B} = -\bar{B} \times \bar{A}$$

Now we may write (1) and (2) as

$$\overline{df} = I\overline{dl} \times \mu'\bar{H} \qquad [3]$$

$$\overline{dH} = \frac{I'\,\overline{dl'} \times \bar{a}_r}{r^2} \qquad [4]$$

$\bar{a}_r$ is the usual unit vector pointing from dl' to dl. $\overline{dH}$ was the contribution to $\bar{H}$ at a point from the element $\overline{dl'}$. Total $\bar{H}$ must be found by summing up vectorially the contribution from all such elements in the system. Although the vector property has been attached quite naturally to $\overline{dl'}$ above, it may as well be given to the current, which has the same direction as $\overline{dl'}$. In subsequent discussions, the vector property will be given to either, depending upon convenience.

$$\bar{H} = \int \frac{\bar{I}'\,dl' \times \bar{a}_r}{r^2} \qquad [5]$$

Equations (3) and (4) define the field strength $\bar{H}$, which expresses the manner in which a current experiences a force and the amount of that force. Equation (5) is often considered the fundamental experimental law. That is to say, it was deduced from experiments on actual systems and serves to correlate measurements on all such systems.

2.19 Electromagnetic System of Units

Current is regarded as a motion of charges; since a unit for electric charge has been defined, there is also defined a unit for current in the electrostatic system. For the study of magnetic fields a new system of units is commonly defined. For this system of units, we shall first set $\mu' = 1$ for vacuum. Two other quantities must now be defined: unit current and unit magnetic field intensity. Equations 2.18(3) and 2.18(5) give two relations between these quantities; force in Eq.2.18(3) is measured in dynes; all distances are in centimeters. Thus, with no constant factors in these equations, the two relations fix units for the two quantities. In terms of complete systems, unit current flowing in a circular loop of wire of 1-cm radius in vacuum produces field strength of 2π at its center (see Art. 2.35); this would then exert a force of 2π dynes per cm of conductor element carrying unit current at right angles to such a field in vacuum.

The above system is known as the electromagnetic system of units (emu). The unit current is called an abampere. The electromagnetic system of units will be used exclusively in all subsequent discussions of magnetic fields in this chapter. Later, as was promised when the electrostatic units were introduced, a single practical system of units will be used to correlate all formulas for application to practical problems. But, it is worth while repeating that we allow the reader some experience with the electromagnetic system of units before restricting our discussions to the practical system, for many valuable texts as well as articles use the former system.

2.20 Vector or Cross Product of Vectors

The vector multiplication defined in Art. 2.18 may be expanded in terms of component vectors as was the scalar product of Art. 2.07. For if $\bar{A}$ and $\bar{B}$ are given in terms of the unit vectors and the components along the three coordinate axes,

$$\bar{A} \times \bar{B} = (A_x\bar{a}_x + A_y\bar{a}_y + A_z\bar{a}_z) \times (B_x\bar{a}_x + B_y\bar{a}_y + B_z\bar{a}_z) \quad [1]$$

From the definition of the vector product and a consideration of the coordinate system, Fig. 2.20, it should be evident that

$$\begin{aligned}
\bar{a}_x \times \bar{a}_y &= \bar{a}_z = -\bar{a}_y \times \bar{a}_x \\
\bar{a}_y \times \bar{a}_z &= \bar{a}_x = -\bar{a}_z \times \bar{a}_y \\
\bar{a}_z \times \bar{a}_x &= \bar{a}_y = -\bar{a}_x \times \bar{a}_z \\
\bar{a}_x \times \bar{a}_x &= 0 = \bar{a}_y \times \bar{a}_y = \bar{a}_z \times \bar{a}_z
\end{aligned}$$

Notice that coordinates were purposely selected so that the sign of the

unit vectors resulting from the product of one unit vector and the succeeding unit vector, in the order xyz, is positive. Such coordinate systems, known as right-handed systems, should always be selected to prevent confusion in signs. To check for a right-handed system, rotate one axis into the succeeding axis in order of writing; a right-hand screw given that motion should then progress in the positive direction along the third axis. Then

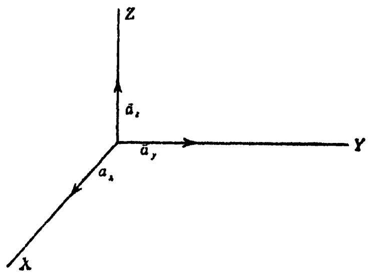


FIG. 2.20. A right-handed system of rectangular coordinates.

$$\bar{A} \times \bar{B} = \bar{a}_x(A_yB_z - A_zB_y) + \bar{a}_y(A_zB_x - A_xB_z) + \bar{a}_z(A_xB_y - A_yB_x) \quad [2]$$

Note that this quantity may also be written as the determinant:

$$\bar{A} \times \bar{B} = \begin{vmatrix} \bar{a}_x & \bar{a}_y & \bar{a}_z \\ A_x & A_y & A_z \\ B_x & B_y & B_z \end{vmatrix} \quad [3]$$

2.21 Line Integral of Magnetic Field; The Curl

We now have available expressions which relate the field distribution to the current distribution, and it might appear that we have generalized sufficiently and can proceed to the more fruitful question of applying these equations to the solution of actual problems. But we are not yet satisfied. There are many other ways of stating these fundamental relations and the ability to get quick answers to electromagnetic problems depends upon being able to choose the best statement of the governing law for that case. Moreover, the understanding of electromagnetic theory is enhanced by one's ability to state it and see it "frontwards, backwards," or if necessary, even "sidewise."

A law which is sometimes given as a fundamental starting point for magnetic fields is that of Biot and Savart, stating that the line integral around any closed path is 4π times the current enclosed.

$$\oint \bar{H} \cdot \overline{dl} = 4\pi I \quad [1]$$

This relation certainly cannot divulge any information not contained in Ampère's law, Eq. 2.18(5), if both are correct, for since field intensity at any point is given in terms of currents from Ampère's law, it could be integrated about any closed path to obtain the line integral in terms of current enclosed. Although both contain the same information, this

integral form of the law will many times be more convenient to use than the previous form, especially when a certain kind of geometrical symmetry exists in a problem. For this reason it will be desirable to derive it from Ampère's law. It is necessary to go through a somewhat rocky mathematical path to show the desired equivalence, and it is not our purpose to engage in rigorous mathematical proofs merely for the exercise. Present interest is rather in concepts which will be valuable when we are confronted with a practical problem. However, a convenient proof of the equivalence between (1) and Eq. 2.18(5) involves the introduction of several concepts which will be useful in later problem solutions, and it is mainly for this reason that the following several articles and problems will be devoted to that proof.

Once before (Art. 2.11) in writing a relation that involved a closed integral, we were able to use the expression to obtain a differential equation. The matter there concerned the relation between the integral of flux diverging from a volume of space and the total charge contained in that volume. The procedure was simply to let the volume shrink to so small a size that it was sufficiently accurate to replace the total charge by the charge density times the volume element. In this way a very handy term called divergence was introduced as the integral of outgoing flux per unit volume. When the divergence was evaluated and equated to 4π times the charge density, a differential equation resulted which actually is capable of serving as a more convenient starting point in many problems than the integral expression from which it was obtained.

Similarly, if we now take (1) and apply it to a very small loop — one so small that it is sufficiently accurate to replace the current linked by the loop by the current density times the small area — it will be possible to derive another extremely useful differential equation and introduce another descriptive vector term. Thus, from (1) we may write

$$\oint \bar{H} \cdot \overline{dl} = 4\pi \bar{\imath} \cdot \overline{\Delta S} \qquad [2]$$

where ΔS is the vanishingly small area and $\bar{\imath}$ is the vector current flow per unit area. In the limit a vector called the curl may be defined by

$$\lim_{\Delta S \to 0} \oint \bar{H} \cdot \overline{dl} = (\text{curl } \bar{H}) \cdot \overline{\Delta S} \qquad [3]$$

Before anything is said about the direction of the curl of a vector and its physical significance, let us go over a very simple example.

Take as the infinitesimal surface a rectangle in cartesian coordinates

parallel to the xy plane (Fig. 2.21).

$$\oint \bar{H} \cdot \overline{dl} = \Delta y\, H_y \big|_{x+\Delta x} - \Delta x\, H_x \big|_{y+\Delta y} - \Delta y\, H_y \big|_x + \Delta x\, H_x \big|_y$$

Now to an accuracy that becomes increasingly better as Δx and Δy go to zero,

$$H_x \bigg|_{y+\Delta y} = H_x \bigg|_y + \Delta y \frac{\partial H_x}{\partial y}\bigg|_y \; ; \quad H_y \bigg|_{x+\Delta x} = H_y \bigg|_x + \Delta x \frac{\partial H_y}{\partial x}\bigg|_x$$

so

$$\oint \bar{H} \cdot \overline{dl} = \left(\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y}\right) \Delta x\, \Delta y \qquad [4]$$

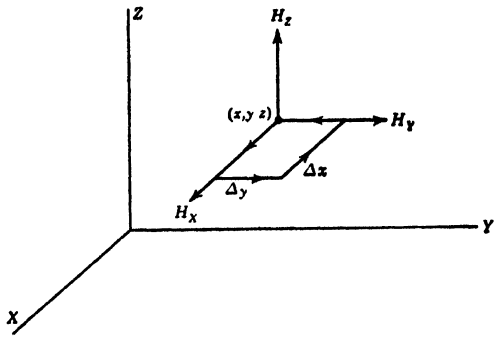


FIG. 2.21.

Now this quantity $(\partial H_y/\partial x) - (\partial H_x/\partial y)$ is a measure of the amount of line integral per unit area lying in a plane perpendicular to the z axis. It tells us how much the magnetic field is curling about a small area in the xy plane where the infinitesimal area vanishes to zero. In a similar way, the curling around infinitesimal areas in the other two planes could be evaluated and would become finally

$$[\text{Curl } \bar{H}]_x = \frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z}$$

$$[\text{Curl } \bar{H}]_y = \frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} \qquad [5]$$

$$[\text{Curl } \bar{H}]_z = \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y}$$

where the subscript denotes the direction of the perpendicular to the plane in which the elementary area lies. This set of equations gives a

differential expression for each component of the curl and clears up the question of direction. Equation (5) may be written also

$$\text{Curl } \bar{H} = \bar{a}_x\left[\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z}\right] + \bar{a}_y\left[\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x}\right] + \bar{a}_z\left[\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y}\right] \quad [6]$$

If the result is compared with the definition of the operator Eq. 2.11(6) and the expansion for the cross product in rectangular coordinates, Eq. 2.20(3), then it is evident that

$$\text{Curl } \bar{H} \equiv \nabla \times \bar{H} \equiv \begin{vmatrix} \bar{a}_x & \bar{a}_y & \bar{a}_z \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ H_x & H_y & H_z \end{vmatrix} \quad [7]$$

This vector operation is read " del cross H " and defined to give the result of (5). From (2), (4), and (5) the following differential equations may be written

$$\lim_{\Delta S \to 0} \int_x \bar{H} \cdot \overline{dl} = \left(\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z}\right) dy\, dz = 4\pi i_x\, dy\, dz$$

$$\lim_{\Delta S \to 0} \int_y \bar{H} \cdot \overline{dl} = \left(\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x}\right) dx\, dz = 4\pi i_y\, dx\, dz \quad [8]$$

$$\lim_{\Delta S \to 0} \int_z \bar{H} \cdot \overline{dl} = \left(\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y}\right) dx\, dy = 4\pi i_z\, dx\, dy$$

Multiplying the first by $\bar{a}_x$, the second by $\bar{a}_y$, etc., and adding, the second and third columns give, from (6),

$$\nabla \times \bar{H} = 4\pi \bar{i} \quad [9]$$

2.22 The Work Integral for Magnetic Fields

When the current density is zero at some point of a magnetic field, Eq. 2.21(9),

$$\nabla \times \bar{H} = 0 \quad [1]$$

Under such circumstances the magnetic field is non-curling or irrotational, just as is true of electrostatic fields, for a quick glance through Art. 2.13 and the definitions of Art. 2.21 should make it evident that

$$\nabla \times \bar{E} = 0 \quad [2]$$

In general, however, for magnetic fields

$$\nabla \times \bar{H} = 4\pi \bar{i}$$

The greater the current density, the greater the curling of the magnetic field, or in the form of Eq. 2.21(1), the greater the total current enclosed in the path, the greater the value of the work integral. This means that in a magnetic field where the current density is not everywhere zero, a motion of conductors carrying the D-C currents (unchanging in time) through some arbitrary paths and back to the original positions will, in general, require the expenditure of energy. Surely, the final field must be the same as the original if currents are maintained constant, and the final and original field energy storage must be equal; the equations show that energy must have been used up just the same. Of course, this does not mean that conservation of energy is violated in this case; the corresponding energy term will appear in induced effects acting on the currents of the system, which effects will be studied in later chapters.

2.23 Vector Magnetic Potential

The curl of a field introduced in Art. 2.21 will prove useful in the development of Eq. 2.21(1) from Ampère's law. This law in vector form, Eq. 2.18(5), gives the magnetic field at point x, y, z. It may also be written

$$\bar{H} = \int \frac{I' \, \overline{dl}' \times \bar{r}}{r^3} \qquad [1]$$

I' is the current in a contributing element $\overline{dl}'$ at point (x', y', z') and $\bar{r}$ is the vector running from dl' to point x, y, z.

$$\bar{r} = \bar{a}_x(x - x') + \bar{a}_y(y - y') + \bar{a}_z(z - z')$$

$$r = \sqrt{(x - x')^2 + (y - y')^2 + (z - z')^2}$$

It may be shown that

$$\frac{\overline{dl}' \times \bar{r}}{r^3} = \text{grad}\left(\frac{1}{r}\right) \times \overline{dl}' \qquad [2]$$

and also that

$$\text{grad}\left(\frac{1}{r}\right) \times \overline{dl}' = \text{curl}\left(\frac{\overline{dl}'}{r}\right) - \frac{1}{r}\,\text{curl}\,\overline{dl}' \qquad [3]$$

But the curl of $\overline{dl}'$ is zero, so that finally

$$\bar{H} = \text{curl} \int I' \left(\frac{\overline{dl}'}{r}\right) = \nabla \times \bar{A} \qquad [4]$$

where

$$\bar{A} \equiv \int I' \frac{\overline{dl}'}{r} \qquad [5]$$

The integral $\bar{A}$ is a vector function of space whose curl gives the magnetic field intensity. It will be called a magnetic potential because when differentiated in a certain way it yields the magnetic field, and more properly the *vector magnetic potential* to distinguish it from other magnetic potentials that can be thought up by analogy with the electrostatic potential and which are not vectors.

In a continuous current distribution throughout a region, of current per unit area $\bar{\imath}$, $\bar{A}$ may be written

$$\bar{A} = \int_V \bar{\imath}\,\frac{dV}{r} \qquad [6]$$

and

$$\bar{H} = \nabla \times \bar{A} \qquad [7]$$

So far as this derivation is concerned, Ampère's law (1) has simply been written in two steps by mathematical maneuvers. That is, (6) and (7) together are the equivalent of (1).

Problem 2.23. Derive Eqs. 2.23(2) and 2.23(3) and explain why it can be said that curl $\bar{dl}'$ is zero.

2.24 Divergence of Magnetic Field

Magnetic field intensity has been written as the curl of a vector, $\bar{A}$. Its divergence is then

$$\nabla \cdot \bar{H} = \nabla \cdot \nabla \times \bar{A} \qquad [1]$$

The result, in rectangular coordinates, is

$$\nabla \cdot \bar{H} = \frac{\partial^2 A_z}{\partial x\,\partial y} - \frac{\partial^2 A_y}{\partial x\,\partial z} + \frac{\partial^2 A_x}{\partial y\,\partial z} - \frac{\partial^2 A_z}{\partial y\,\partial x} + \frac{\partial^2 A_y}{\partial z\,\partial x} - \frac{\partial^2 A_x}{\partial z\,\partial y} \equiv 0 \qquad [2]$$

since

$$\frac{\partial^2}{\partial x\,\partial y} = \frac{\partial^2}{\partial y\,\partial x}, \quad \text{etc.}$$

Notice that the evaluation of the divergence of the curl of $\bar{A}$ was independent of the value of $\bar{A}$, so then the divergence of the curl of any vector is identically zero.

A major difference between electric and magnetic fields is here apparent, for unlike the electric field, the magnetic field must have zero divergence everywhere. That is, when the magnetic field is due to currents, there are no sources of magnetic flux which correspond to the electric charges as sources of electric flux. Fields with zero divergence such as these are consequently often called source free fields.

Magnetic field concepts are often developed from an exact parallel with electric fields by considering the concept of isolated magnetic poles as sources of magnetic flux, corresponding to the charges of electrostatics. The result of zero divergence still seems entirely applicable since such poles have never been isolated, but seem to appear in nature as equal and opposite pairs. In other words, it is correct to write

$$\nabla \cdot \bar{B} = 0$$

where $\bar{B}$ is called the magnetic flux density and

$$\bar{B} = \mu' \bar{H} \qquad [3]$$

2.25 Differential Equations for Vector Magnetic Potential

We are now in a position to derive the expression for curl $\bar{H}$ in terms of the currents of the system, which was written without proof in Art. 2.21. To do so, let us make good use of the vector magnetic potential.

$$\nabla \times \bar{H} = \nabla \times \nabla \times \bar{A} \qquad [1]$$

from Eq. 2.23(7). The identity

$$\nabla \times \nabla \times \bar{A} = -\nabla^2 \bar{A} + \nabla(\nabla \cdot \bar{A}) \qquad [2]$$

will often be useful; it can be easily verified by the definitions given so far and the further definition for rectangular coordinates,

$$\nabla^2 \bar{A} = \bar{a}_x \nabla^2 A_x + \bar{a}_y \nabla^2 A_y + \bar{a}_z \nabla^2 A_z \qquad [3]$$

It may be proved that if $\bar{A}$ is defined by Eq. 2.23(6),

$$\nabla \cdot \bar{A} = 0 \qquad [4]$$

There remains then only

$$\nabla \times \bar{H} = -\nabla^2 \bar{A} \qquad [5]$$

Note that from the definition of $\bar{A}$, Eq. 2.23(6)

$$A_x = \int_V \frac{i_x\, dV}{r} \qquad [6]$$

This should be compared with Poisson's equation and the integral expression for potential in electrostatics

$$\Phi = \int_V \frac{\rho\, dV}{\epsilon' r} \qquad \nabla^2 \Phi = -\frac{4\pi\rho}{\epsilon'} \qquad [7]$$

Although these equations were obtained from a consideration of the properties of electrostatic fields, the first of these two equations (7) may be considered as the solution in integral form of the second, for any scalar functions Φ and ρ/ϵ'. Consequently by direct analogy between

(6) and (7), we write

$$\nabla^2 A_x = -4\pi i_x \tag{8}$$

Similarly for the other components, so that by making use of (3),

$$\nabla^2 \bar{A} = -4\pi(\bar{a}_x i_x + \bar{a}_y i_y + \bar{a}_z i_z) = -4\pi \bar{\imath} \tag{9}$$

Finally, from (5)

$$\text{Curl } \bar{H} \equiv \nabla \times \bar{H} = 4\pi \bar{\imath} \tag{10}$$

Problem 2.25(*a*). Prove that Eq. 2.25(2) is correct.

Problem 2.25(*b*). Remembering Eq 2.23(6), show that

$$\nabla \cdot \bar{A} = -\int_V \bar{\imath} \cdot \text{grad}_{x',y',z'} \left(\frac{1}{r}\right) dV$$

where x, y, z, is the point at which $\bar{A}$ and $\bar{H}$ are being studied and x', y', z' is the point of location of the current density $\bar{\imath}$.

Problem 2.25(*c*). Show that

$$\nabla \cdot (S\bar{p}) \equiv S\nabla \cdot \bar{p} + \bar{p} \cdot \nabla S$$

Problem 2.25(*d*). Using the results of the two previous problems, show that Eq. 2.25(4) is correct.

2.26 Stokes' Theorem

Just as the divergence should be thought of as a flux flow per unit volume, the curl should be thought of as a line integral per unit area, at a point in space. Just as the divergence theorem (Art. 2.12) states that the total flux flow out of any volume may be obtained by integration of the divergence throughout that volume, there is another theorem which states that the line integral around any surface may be obtained by integrating the normal components of the curl over that surface. If the surface is broken up into a large number of infinitesimal areas as shown in Fig. 2.26, it is known from the definition of curl that for each of these infinitesimal areas

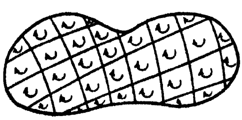


FIG. 2.26.

$$\oint \bar{H} \cdot \overline{dl} = \text{curl } \bar{H} \cdot dS$$

If contributions from infinitesimal areas are summed over all the surface, the line integral must disappear for all internal areas, since a boundary is first traversed in one direction and then later in the opposite direction in determining the contribution from an adjacent area.

The only places where these contributions do not disappear are along the outer boundary, so that the result of the summation is then the line integral of the vector around the boundary.

$$\oint \bar{H} \cdot \overline{dl} = \int_S \text{curl } \bar{H} \cdot \overline{dS} \equiv \int_S \nabla \times \bar{H} \cdot \overline{dS} \qquad [1]$$

This relation is known as Stokes' theorem, and as with the divergence theorem, holds for any vector field.

2.27 Derivation of Integral Law of Biot and Savart

From Stokes' theorem, the line integral of magnetic field around any path may be obtained. Combining Eqs. 2.25 (10) and 2.26(1),

$$\oint \bar{H} \cdot \overline{dl} = \int_S (4\pi\bar{\imath}) \cdot \overline{dS} = 4\pi I \qquad [1]$$

I is the current enclosed by the path. This is the equation stated in Art. 2.21 without proof. We have now shown that it follows from Ampère's law, Eq. 2.18(5), and in doing so, have introduced a vector potential for magnetic fields which will be useful in subsequent field problems.

2.28 Scalar Potentials for Magnetic Fields

Suppose we had stubbornly attempted to derive the magnetic field as the gradient of a scalar potential, as was done for the electrostatic field, say,

$$\bar{H} = \text{grad } \Phi_m$$

The curl of $\bar{H}$ from Eq. 2.21(7) would then be

$$\nabla \times \bar{H} = \begin{vmatrix} \bar{a}_x & \bar{a}_y & \bar{a}_z \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ \dfrac{\partial \Phi_m}{\partial x} & \dfrac{\partial \Phi_m}{\partial y} & \dfrac{\partial \Phi_m}{\partial z} \end{vmatrix}$$

Now if this is expanded it will be found that it is identically zero [because of the cancellation of terms like $(\partial^2\Phi_m/\partial x\partial y) - (\partial^2\Phi_m/\partial y\partial x)$]. Since the curl of the gradient is identically zero for any vector field, we cannot hope to specify a field which does not have zero curl as the gradient of a scalar potential. When a field is properly expressible as the gradient of a scalar potential, the line integral of that field between any two points is independent of the path, and the line integral about any closed path is zero.

Of course, it is true that in a current-free region $\nabla \times \bar{H} = 0$ and $\bar{H}$ may be derived as the gradient of a scalar potential throughout such a region. This magnetic scalar potential will satisfy Laplace's equation, just as would the electrostatic potential in a region containing no charges. Such a potential is useful whenever one is interested in a region containing no currents, given the conditions at all boundaries of that region. The vector magnetic potential, valid as it is for either current-free or current-carrying regions, has more general usefulness.

2.29 Uniqueness of the Vector Magnetic Potential

In electrostatics there were any number of scalar potential functions whose gradients gave the same electric field; all these potentials differed by a constant which, having zero gradient, could change the potential without affecting the field. In a similar but perhaps a bit more complex way, it should be clear that there are ever so many vector potentials whose curls will be nevertheless all the same. All these vector potentials will differ by the addition of some function whose curl is zero. For example, we found in Art. 2.28 that the curl of the gradient of any scalar always turns out to be zero no matter what scalar function is considered. Thus if $\bar{A}$ is a vector whose curl is equal to the magnetic field intensity $\bar{H}$, then $\bar{A} + \nabla\psi$ (where ψ is any scalar function) will also be such a vector. It appears that, just as the arbitrary constant in the scalar electric potential was chosen as a matter of convenience, so might $\nabla\psi$ be chosen to arrive at the most convenient function for the vector magnetic potential.

We are quite accustomed to handling the constant in the scalar electric potential with ease, to make one or another conductor the zero potential electrode and to find the fields, without being held up by worry over the question of uniqueness. Whenever the vector magnetic potential is used later in this text, the conditions necessary to fix on one of the many functions having the same curl will be considered further. For instance, in defining Φ and $\bar{A}$ from

$$\Phi = \int_V \frac{\rho dV}{\epsilon' r}$$

$$\bar{A} = \int_V \frac{\bar{\imath} dV}{r}$$

the volume is all space in each, and the charge and current distributions are assumed to be known and definite. Evidently in this case there are no questions of uniqueness, arbitrary constants, or gradients of scalars — the functions Φ and $\bar{A}$ are definitely determined. But this only means

that we have automatically chosen to fix these arbitrary factors. We might, for example, have written

$$\Phi = \int_V \frac{\rho\, dV}{\epsilon' r} + C$$

$$\bar{A} = \int_V \frac{\bar{\imath}\, dV}{r} + \nabla\psi$$

and the magnetic and electric fields would have been the same. It is important to remember that though the potentials are of use in simplifying the mathematics or often the whole approach to many problems in electromagnetics, their successful use comes best after the "feel" for their properties is obtained by working many examples. This we shall endeavor to make possible during later parts of the text.

SIMPLE APPLICATIONS OF THE THEORY

2.30 Field about a Charged Cylindrical Conductor

It is important that as much unnecessary work as possible be eliminated by consideration of geometrical symmetry whenever it exists in physical problems. In an infinitely long conductor of circular cross section charged uniformly with charge q per unit length (Fig. 2.30) symmetry requires that the electric field must be entirely radial and unvarying with angle. Gauss's law requires that the flux passing through an imaginary cylindrical surface at any radius r be $4\pi q$ per unit length, so the flux per unit area at r becomes

$$D_r = \frac{4\pi q}{2\pi r} = \frac{2q}{r} \qquad [1]$$

and

$$E_r = \frac{2q}{\epsilon' r} \qquad [2]$$

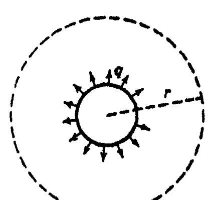


Fig. 2.30. Charged cylinder.

Notice that this result is independent of the diameter of the charged cylinder.

2.31 Boundary between Two Dielectrics

The boundary between two dielectrics may be investigated with the aid of Gauss's law and the requirement that the electric field integral about any closed path be zero. If there is no charge on the boundary, an imaginary small surface as indicated by the dotted line of Fig. 2.31 must enclose no charge. If subscript n denotes components normal to the

surface of area ΔS,

$$\Delta S\, D_{1n} + \Delta S(-D_{2n}) = 0 \qquad [1]$$

or

$$D_{1n} = D_{2n}$$
$$\epsilon_1' E_{1n} = \epsilon_2' E_{2n} \qquad [2]$$

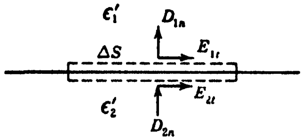


FIG. 2.31. Cross section, showing surface separating two dielectrics.

If a charge q is taken around a closed path having Δl, half its total length, on each side of the boundary, the work done is

$$-\Delta l\, qE_{1t} + \Delta l\, qE_{2t} = 0$$

or

$$E_{1t} = E_{2t} \qquad [3]$$

The subscript t denotes components tangential to the surface.

Thus the conditions at the boundary between two dielectrics are: normal components of displacement and tangential components of electric intensity must be equal on the two sides of the boundary; in other words, both are continuous. In general, then, the direction of $\bar{D}$ or $\bar{E}$ will change in crossing the boundary between dielectrics of different dielectric constant.

Problem 2.31. If the field vector makes an angle θ_1 with the normal in region 1 of the above example, what angle does it have in region 2?

2.32 The Dipole

A study of the field due to a dipole, a pair of equal but opposite charges separated by a very small distance, will be of interest in later work on radiation.

By definition, the electric moment $\overline{m}$ of a dipole is a vector whose magnitude is given by the product of one of the charges and the distance between the two, and whose direction is given by the direction of the line drawn from negative to positive charge. If the dipole is as shown in Fig. 2.32, the potential at P is the sum of contributions from the two charges.

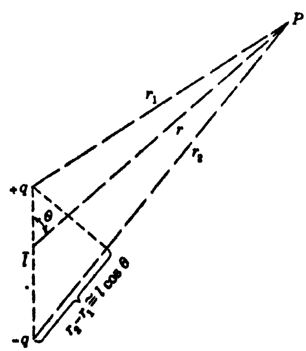


FIG. 2.32. Dipole and distant point P.

$$\Phi = \frac{q}{\epsilon'}\left(\frac{1}{r_1} - \frac{1}{r_2}\right) \qquad [1]$$

And, if the distance l is small compared with r, this is

$$\Phi = \frac{ql \cos \theta}{\epsilon' r^2} = \frac{m \cos \theta}{\epsilon' r^2}$$

$$= \frac{\overline{m} \cdot \bar{r}}{\epsilon' r^3} \qquad [2]$$

The electric field due to the dipole is

$$\bar{E} = -\nabla\Phi$$

In Art. 2.38 an expression for gradient in spherical coordinates is given.

$$\nabla\Phi = \bar{a}_r\left(\frac{\partial\Phi}{\partial r}\right) + \bar{a}_\theta\left(\frac{1}{r}\frac{\partial\Phi}{\partial\theta}\right) + \bar{a}_\phi\left(\frac{1}{r\sin\theta}\frac{\partial\Phi}{\partial\phi}\right)$$

Then if the dipole is parallel to the polar axis of spherical coordinates

$$E_r = -\frac{\partial\Phi}{\partial r} = \frac{2m\cos\theta}{\epsilon' r^3}$$

$$E_\theta = -\frac{1}{r}\frac{\partial\Phi}{\partial\theta} = \frac{m\sin\theta}{\epsilon' r^3} \qquad [3]$$

$$E_\phi = \frac{1}{r\sin\theta}\frac{\partial\Phi}{\partial\phi} = 0$$

Problem 2.32(*a*). Extend the concept to a shell or cap of thickness t with a constant distribution of charge density $+\rho_s$ on one side and $-\rho_s$ on the other, showing that

$$\Phi = \frac{M\Omega}{\epsilon'}$$

where Ω is the solid angle subtended by the entire shell at point P, and M is defined as $\rho_s t$.

Problem 2.32(*b*). Find the field from an axial quadripole formed by two dipoles of the same moments but opposite sign removed from each other by a distance l_2 in the direction of the dipole moments.

Problem 2.32(*c*). Repeat for a quadripole formed by separating the equal and opposite dipoles by a distance l_2 normal to the dipole moments.

2.33 Energy of an Electrostatic System

The work required to move a charge in the vicinity of a system of charges was discussed in the study of the electrostatic potential. The work done must appear as energy stored in the system, and consequently the potential energy of a system of charges may be computed from the amount and position of the charges. If a charge q' is brought from

infinity to a point at distance r from charge q, the work done was shown to be

$$U_E = \frac{qq'}{\epsilon' r}$$

Then for a large number of charges

$$U_E = \frac{1}{2}\sum_m \sum_n \frac{q_n q_m}{\epsilon' r_{nm}} \qquad [1]$$

The factor ½ appears since n and m are each summed over all the particles, and by this convention each contribution of energy is included twice.

In (1) it is apparent that the term for which $q_n = q_m$ will cause difficulty. It is the energy of an isolated point charge, and the value of r_{nm} is zero. This says that the energy required to locate any finite amount of charge at a point is infinite. Such a conclusion is not incorrect; rather, it is an expected result since to build up charge at a point involves infinite repelling forces between the additional charge being introduced and the amount already there. Actually (and, in fact, almost for this very reason) we do not have charges concentrated at points; instead, there always is a certain amount of space distribution. Recognition of this suggests that an expression for energy more useful than (1) may be obtained.

If it is noted from Eq. 2.14(2) that the potential at the mth charge is

$$\Phi_m = \sum_n \frac{q_n}{\epsilon' r_{nm}}$$

then (1) may be written

$$U_E = \tfrac{1}{2}\sum_n \Phi_m q_m \qquad [2]$$

Or, extending to a system with continuously varying charge density ρ per unit volume

$$U_E = \tfrac{1}{2}\int_V \rho\Phi \, dV \qquad [3]$$

This expression may be altered to

$$U_E = -\frac{1}{8\pi}\int_V \bar{D} \cdot \operatorname{grad} \Phi \, dV = \frac{1}{8\pi}\int_V \bar{D} \cdot \bar{E} \, dV \qquad [4]$$

This result seems to say that the energy is actually in the electric field, each element of volume dV appearing to contain the amount of energy

$$dU_E = \frac{1}{8\pi} \bar{D} \cdot \bar{E} \, dV \qquad [5]$$

The right answer is obtained if this "energy density" picture is used. Actually, we know only that the total energy stored in the system will be correctly computed by the total integral of (4).

It is interesting to check these results against a case with which we are already familiar. Consider a parallel plate condenser of capacity C and a voltage between plates of V. The energy is known to be $\frac{1}{2}CV^2$ which is commonly obtained by integrating the product of instantaneous current and instantaneous voltage over the time of charging. The result may also be obtained by integrating the energy distribution in the field throughout the volume between plates according to the concepts of (4) and (5). For plates of area A closely spaced so that the end effects may be neglected, the magnitude of field at every point in the dielectric is $E = V/d$ (d = distance between plates).

Hence

$$D = \frac{\epsilon' V}{d}$$

$$\text{Stored energy} = \frac{1}{8\pi} Ad \left(\frac{\epsilon' V}{d}\right)\left(\frac{V}{d}\right)$$

$$= \frac{1}{2}\left(\frac{\epsilon' A}{4\pi d}\right) V^2 = \frac{1}{2} CV^2 \qquad [6]$$

Problem 2.33. Remembering that div $\bar{D} = 4\pi\rho$, derive Eq. 2.33(4) from 2.33(3). (*Hint.* In order to evaluate certain surface integrals which may appear, consider the surface at infinity, since this is a surface including all the energy.)

2.34 Energy of a Magnetostatic System

It is possible to derive an expression (this is done in several of the references) similar to Eq. 2.33(5) for the energy stored in the magnetic field:

$$U_H = \frac{1}{8\pi}\int_V \bar{B} \cdot \bar{H}\, dV \qquad [1]$$

It is useful to note that when the field is entirely due to static currents and the permeability of the medium is constant, the energy storage may also be expressed in rather simple fashion in terms of the magnitude and distribution of the currents. For the simple case of a magnetic field due to current I in a single circuit,

$$U_H = \tfrac{1}{2}LI^2 \qquad [2]$$

where L is a constant, the inductance of the circuit, that depends upon the space distribution of the current. We postpone a discussion of this

constant until later chapters, but the student should, by consulting the references, satisfy himself that (2) is valid for the situation described.

2.35 Magnetic Field at the Center of a Circular Current Loop

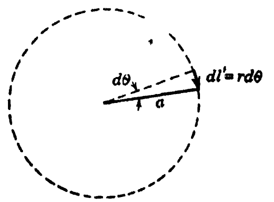


FIG. 2.35. Circular loop carrying current.

Ampère's law may be used in finding the field at the center of a circular current loop.

$$d\bar{H} = \frac{I\bar{dl}' \times \bar{r}}{r^3}$$

In this case simplification arises since r is always constant (equal to a, the radius of the loop), $\bar{dl}'$ and $\bar{r}$ are perpendicular so that the cross product reduces to rdl', and all the contributions $d\bar{H}$ have a common direction normal to the plane of the loop. Thus a scalar sum may be substituted for the vector sum.

$$dH = \frac{I(ad\theta)}{a^2}$$

$$H = \frac{I}{a}\int_0^{2\pi} d\theta = \frac{2\pi I}{a} \qquad [1]$$

2.36 Magnetic Field of a Straight Current

This case may also be solved from Ampère's law, but symmetry permits the use of the integral law of Biot and Savart to obtain the answer at once. The line integral of $\bar{H}$ about any closed path surrounding the wire is $4\pi I$. Symmetry requires that $\bar{H}$ have only a tangential component H_ϕ and no variations with ϕ. So if the path of integration is taken as a circle at radius r from the center of the wire,

$$H_\phi = \frac{4\pi I}{2\pi r} = \frac{2I}{r} \qquad [1]$$

Problem 2.36. Find the field at any point inside an infinitely long solenoid having n turns per centimeter, in terms of solenoid current I.

OFTEN-USED RELATIONS IN OFTEN-USED COORDINATE SYSTEMS

2.37 Orthogonal Curvilinear Coordinates

When it was necessary to refer to a coordinate system in the preceding articles, rectangular coordinates were used exclusively. There was actually no loss in generality in the results obtained, since if it had been necessary to transfer at any point to a new coordinate system, it could

have been done by an ordinary transformation of variables. However, if it had been desirable from the start to carry through some other coordinate system, we would merely have defined the expressions for scalar product, vector product, gradient, divergence, and curl in terms of that coordinate system. To eliminate either the transformation of variables or the reconstruction from the beginning each time a new coordinate system is considered, it will be more convenient to obtain general expressions for the defined vector operations which give the form of these operations in rectangular coordinates, cylindrical coordinates, spherical coordinates, or in any system which is an orthogonal system.

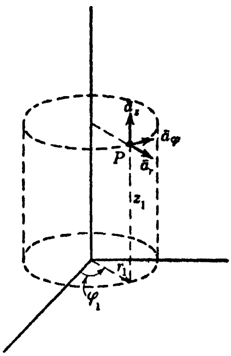


FIG. 2.37*a*. System of circular cylindrical coordinates.

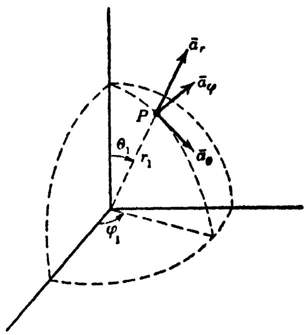


FIG. 237*b*. System of spherical coordinates.

The intersection of two surfaces is a line; the intersection of three surfaces is a point; thus the coordinates of a point are usually given as the three parameters referring to three sets of surfaces such that the parameter attached to a particular surface is constant along that surface. If the lines of intersection of these surfaces are at right angles, the system is said to be orthogonal. In this book we shall use only rectangular, cylindrical, and spherical coordinates, all of which are orthogonal. In rectangular coordinates the three planes $x = x_1$, $y = y_1$, $z = z_1$ intersect at a point which is designated by the coordinates x_1, y_1, z_1. In cylindrical coordinates (only circular cylinders being considered here) the surfaces are (*a*) a set of circular cylinders (r = constant), (*b*) a set of planes all passing through the axis (ϕ = constant), (*c*) a set of planes normal to the axis (z = constant). Coordinates of a point are then given by r_1, ϕ_1, z_1 (Fig. 2.37*a*).

In spherical coordinates the surfaces are (*a*) a set of spheres (r = constant), (*b*) a set of circular cones about the axis (θ = constant),

(c) a set of planes all passing through the polar axis (ϕ = constant). The intersection of sphere $r = r_1$, circular cone $\theta = \theta_1$, and plane through the polar axis $\phi = \phi_1$ gives a point whose coordinates are said to be (r_1, θ_1, ϕ_1) as shown in Fig. 2.37b. All the coordinate systems drawn are right-handed systems (see Art. 2.20).

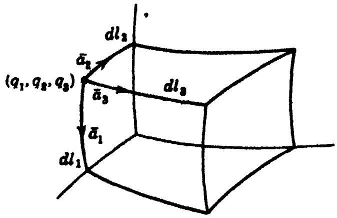


FIG. 2.37c Element in arbitrary orthogonal curvilinear coordinates.

Suppose a point in space is thus defined in any orthogonal system by the coordinate surfaces q_1, q_2, q_3. These then intersect at right angles and a set of three unit vectors, $\bar{a}_1$, $\bar{a}_2$, $\bar{a}_3$, may be placed at this point. These should point in the direction of increasing coordinates. (Fig. 2.37c.) The three coordinates need not necessarily express directly a distance (consider, for example, the angles of spherical coordinates) so that the differential elements of distance must be expressed:

$$dl_1 = h_1\, dq_1, \quad dl_2 = h_2\, dq_2, \quad dl_3 = h_3\, dq_3 \tag{1}$$

where h_1, h_2, h_3 in the most general case may each be functions of all three coordinates, q_1, q_2, q_3.

Scalar and Vector Products. A reference to the fundamental definitions of the two vector multiplications will show that these do not change in form in orthogonal curvilinear coordinates. Thus, for scalar or dot product

$$\bar{A} \cdot \bar{B} = A_1B_1 + A_2B_2 + A_3B_3 \tag{2}$$

and for the vector or cross product

$$\bar{A} \times \bar{B} = \begin{vmatrix} \bar{a}_1 & \bar{a}_2 & \bar{a}_3 \\ A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \end{vmatrix} \tag{3}$$

When one of these vectors is replaced by the operator ∇, the above expressions do not hold, as will be shown below.

Gradient. According to previous definitions, the gradient of any scalar Φ will be a vector whose component in any direction is given by the change of Φ for a change in distance along that direction. Thus

$$\nabla\Phi = \bar{a}_1 \frac{\partial\Phi}{h_1\, \partial q_1} + \bar{a}_2 \frac{\partial\Phi}{h_2\, \partial q_2} + \bar{a}_3 \frac{\partial\Phi}{h_3\, \partial q_3} \tag{4}$$

Divergence. Divergence requires the evaluation of the outgoing flux per unit area for a small volume. As in Art. 2.11, a small volume element is chosen, having its six sides lying in the surfaces

$$q_1 - \frac{dq_1}{2}; \quad q_1 + \frac{dq_1}{2}; \quad q_2 - \frac{dq_2}{2}; \quad q_2 + \frac{dq_2}{2}; \quad q_3 - \frac{dq_3}{2}; \quad q_3 + \frac{dq_3}{2}$$

The volume of the element is then $h_1h_2h_3\,dq_1\,dq_2\,dq_3$. If $\bar{D}$ is the value of displacement vector at the center of the element, it may be written $\bar{D} = \bar{a}_1D_1 + \bar{a}_2D_2 + \bar{a}_3D_3$. The flux passing through any elemental area perpendicular to D_1 at q_1, q_2, q_3 is then $d\psi_1\big|_{q_1} = D_1(h_2h_3\,dq_2\,dq_3)$.

For general curvilinear coordinates, not only the field components but also the area of cross section of an elemental volume may vary with the coordinate. So, now

$$d\psi_1\Big|_{q_1-\frac{dq_1}{2}} = D_1h_2h_3\,dq_2\,dq_3 - \frac{dq_1}{2}\frac{\partial}{\partial q_1}(D_1h_2h_3\,dq_2\,dq_3)$$

$$d\psi_1\Big|_{q_1+\frac{dq_1}{2}} = D_1h_2h_3\,dq_2\,dq_3 + \frac{dq_1}{2}\frac{\partial}{\partial q_1}(D_1h_2h_3\,dq_2\,dq_3)$$

The net outgoing flux for these two sides is thus

$$d\psi_1\Big|_{q_1+\frac{dq_1}{2}} - d\psi_1\Big|_{q_1-\frac{dq_1}{2}} = \frac{\partial}{\partial q_1}(D_1h_2h_3\,dq_2\,dq_3)\,dq_1$$

Similarly, for the other two directions, so that finally $\nabla \cdot \bar{D}$ = net flux flow divided by the volume, is expressed by

$$\nabla \cdot \bar{D} = \frac{1}{h_1h_2h_3\,dq_2\,dq_3}\left[\frac{\partial(h_2h_3\,dq_2\,dq_3\,D_1)}{\partial q_1} + \cdots\right]$$

or

$$\nabla \cdot \bar{D} = \frac{1}{h_1h_2h_3}\left[\frac{\partial(h_2h_3D_1)}{\partial q_1} + \frac{\partial(h_1h_3D_2)}{\partial q_2} + \frac{\partial(h_2h_1D_3)}{\partial q_3}\right] \qquad [5]$$

As an example, consider the case of spherical coordinates

$$\begin{aligned} dl_1 &= dr \\ dl_2 &= r\,d\theta \\ dl_3 &= r\sin\theta\,d\phi \end{aligned} \qquad \text{so} \qquad \begin{cases} h_1 = 1 \\ h_2 = r \\ h_3 = r\sin\theta \end{cases} \qquad [6]$$

and

$$\nabla \cdot \bar{D} = \frac{1}{r^2\sin\theta}\left[\frac{\partial}{\partial r}(r^2\sin\theta\,D_r) + \frac{\partial}{\partial\theta}(r\sin\theta\,D_\theta) + \frac{\partial}{\partial\phi}(r\,D_\phi)\right]$$

$$= \frac{1}{r^2}\frac{\partial}{\partial r}(r^2D_r) + \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(\sin\theta\,D_\theta) + \frac{1}{r\sin\theta}\frac{\partial D_\phi}{\partial\phi}$$

Curl. Just as the development of the divergence in rectangular coordinates from Art. 2.11 was generalized above to fit any general orthogonal system, so does the revision of the curl from Art. 2.21 follow in an exactly similar fashion to give

$$\nabla \times \bar{H} = \begin{vmatrix} \dfrac{\bar{a}_1}{h_2 h_3} & \dfrac{\bar{a}_2}{h_3 h_1} & \dfrac{\bar{a}_3}{h_1 h_2} \\ \dfrac{\partial}{\partial q_1} & \dfrac{\partial}{\partial q_2} & \dfrac{\partial}{\partial q_3} \\ h_1 H_1 & h_2 H_2 & h_3 H_3 \end{vmatrix} \qquad [7]$$

Divergence of the Gradient. By combining the general expressions for gradient and divergence (4) and (5)

$$\nabla \cdot \nabla \Phi = \nabla^2 \Phi$$

$$= \frac{1}{h_1 h_2 h_3}\left[\frac{\partial}{\partial q_1}\left(\frac{h_2 h_3}{h_1}\frac{\partial \Phi}{\partial q_1}\right) + \frac{\partial}{\partial q_2}\left(\frac{h_3 h_1}{h_2}\frac{\partial \Phi}{\partial q_2}\right) + \frac{\partial}{\partial q_3}\left(\frac{h_1 h_2}{h_3}\frac{\partial \Phi}{\partial q_3}\right)\right] \qquad [8]$$

∇^2 *of Vectors.* For expressions such as $\nabla^2 \bar{A}$ in curvilinear coordinates, where differentiation is to be performed on a vector, it is necessary to consider a definition more general than that of Art. 2.25 (unless as in that article we use only rectangular coordinates). By definition,

$$\nabla^2 \bar{A} = \nabla \cdot (\nabla \bar{A}) \qquad [9]$$

where

$$\nabla \bar{A} = \bar{a}_1 \frac{\partial \bar{A}}{h_1\, \partial q_1} + \bar{a}_2 \frac{\partial \bar{A}}{h_2\, \partial q_2} + \bar{a}_3 \frac{\partial \bar{A}}{h_3\, \partial q_3}$$

For example, in spherical coordinates

$$\nabla \bar{A} = \bar{a}_r \frac{\partial \bar{A}}{\partial r} + \frac{\bar{a}_\theta}{r}\frac{\partial \bar{A}}{\partial \theta} + \frac{\bar{a}_\phi}{r \sin \theta}\frac{\partial \bar{A}}{\partial \phi}$$

There appear such terms as

$$\frac{\partial \bar{A}}{\partial r} = \frac{\partial}{\partial r}\left[\bar{a}_r A_r + \bar{a}_\theta A_\theta + \bar{a}_\phi A_\phi\right]$$

$$= \bar{a}_r \frac{\partial A_r}{\partial r} + A_r \frac{\partial \bar{a}_r}{\partial r} + \bar{a}_\theta \frac{\partial A_\theta}{\partial r} + A_\theta \frac{\partial \bar{a}_\theta}{\partial r} + \bar{a}_\phi \frac{\partial A_\phi}{\partial r} + A_\phi \frac{\partial \bar{a}_\phi}{\partial r} \qquad [10]$$

We cannot neglect the variation of the unit vectors $\bar{a}_r$, $\bar{a}_\theta$, $\bar{a}_\phi$ with r, θ, ϕ, for it is evident that although the magnitudes always remain unity, the directions of the unit vectors may change with changes in any of the coordinates. (Such a situation is not true of the simpler, rectangular coordinate unit vectors.)

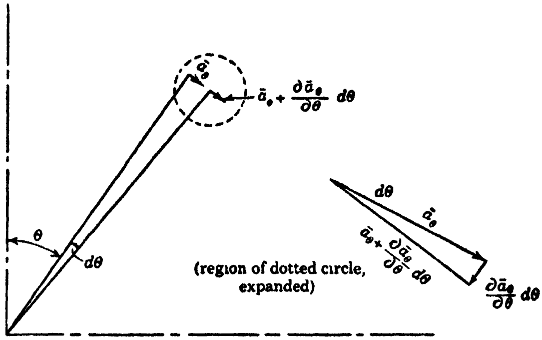


FIG. 2.37*d*.

In spherical coordinates, note that there is no change in direction of any of the three unit vectors if r alone is changed, but consider a small change in θ and its effect on $\bar{a}_\theta$. The vector $\dfrac{\partial \bar{a}_\theta}{\partial \theta}\, d\theta$ from Fig. 2.37*d* is seen to have a magnitude of $d\theta$ and a direction given by $-\bar{a}_r$. Thus $\partial \bar{a}_\theta/\partial \theta = -a_r$.

Problem 2.37. (*a*) From studies such as that of Fig. 2.37*d*, find all other partials of unit vectors in Eq. 2.37(10) and complete $\nabla^2 \bar{A}$ for spherical coordinates.

(*b*) Find $\nabla^2 \bar{A}$ for cylindrical coordinates.

2.38 Summary of Useful Vector Relations

From the general equations for all orthogonal systems, the forms of the vector operations in these three most common systems are found to be as follows.

Rectangular Coordinates

$$q_1 = x, \qquad q_2 = y, \qquad q_3 = z$$

$$h_1 = 1, \qquad h_2 = 1, \qquad h_3 = 1$$

$$\nabla \Phi = \bar{a}_x \frac{\partial \Phi}{\partial x} + \bar{a}_y \frac{\partial \Phi}{\partial y} + \bar{a}_z \frac{\partial \Phi}{\partial z}$$

$$\nabla \cdot \bar{D} = \frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} + \frac{\partial D_z}{\partial z}$$

$$\nabla \times \bar{H} = \bar{a}_x \left(\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} \right) + \bar{a}_y \left(\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} \right) + \bar{a}_z \left(\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} \right)$$

$$\nabla^2 \Phi = \frac{\partial^2 \Phi}{\partial x^2} + \frac{\partial^2 \Phi}{\partial y^2} + \frac{\partial^2 \Phi}{\partial z^2}$$

All partials of unit vectors $\left(\frac{\partial \bar{a}_x}{\partial x}, \frac{\partial \bar{a}_x}{\partial y}, \text{etc.} \right)$ are zero.

Cylindrical Coordinates

$$q_1 = r, \qquad q_2 = \phi, \qquad q_3 = z$$

$$h_1 = 1, \qquad h_2 = r, \qquad h_3 = 1$$

$$\nabla \Phi = \bar{a}_r \frac{\partial \Phi}{\partial r} + \bar{a}_\phi \frac{1}{r} \frac{\partial \Phi}{\partial \phi} + \bar{a}_z \frac{\partial \Phi}{\partial z}$$

$$\nabla \cdot \bar{D} = \frac{1}{r} \frac{\partial}{\partial r} (r D_r) + \frac{1}{r} \frac{\partial D_\phi}{\partial \phi} + \frac{\partial D_z}{\partial z}$$

$$\nabla \times \bar{H} = \bar{a}_r \left[\frac{1}{r} \frac{\partial H_z}{\partial \phi} - \frac{\partial H_\phi}{\partial z} \right] + \bar{a}_\phi \left[\frac{\partial H_r}{\partial z} - \frac{\partial H_z}{\partial r} \right] + \bar{a}_z \left[\frac{1}{r} \frac{\partial (r H_\phi)}{\partial r} - \frac{1}{r} \frac{\partial H_r}{\partial \phi} \right]$$

$$\nabla^2 \Phi = \frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial \Phi}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2 \Phi}{\partial \phi^2} + \frac{\partial^2 \Phi}{\partial z^2}$$

All partials of unit vectors are zero except

$$\frac{\partial \bar{a}_r}{\partial \phi} = \bar{a}_\phi \qquad \frac{\partial \bar{a}_\phi}{\partial \phi} = -\bar{a}_r$$

Spherical Coordinates

$$\nabla \Phi = \bar{a}_r \frac{\partial \Phi}{\partial r} + \bar{a}_\theta \frac{1}{r} \frac{\partial \Phi}{\partial \theta} + \frac{\bar{a}_\phi}{r \sin \theta} \frac{\partial \Phi}{\partial \phi}$$

$$\nabla \cdot \bar{D} = \frac{1}{r^2} \frac{\partial}{\partial r} (r^2 D_r) + \frac{1}{r \sin \theta} \frac{\partial}{\partial \theta} (\sin \theta D_\theta) + \frac{1}{r \sin \theta} \frac{\partial D_\phi}{\partial \phi}$$

$$\nabla \times \bar{H} = \frac{\bar{a}_r}{r \sin \theta} \left[\frac{\partial}{\partial \theta} (H_\phi \sin \theta) - \frac{\partial H_\theta}{\partial \phi} \right] + \frac{\bar{a}_\theta}{r} \left[\frac{1}{\sin \theta} \frac{\partial H_r}{\partial \phi} - \frac{\partial}{\partial r} (r H_\phi) \right] + \frac{\bar{a}_\phi}{r} \left[\frac{\partial}{\partial r} (r H_\theta) - \frac{\partial H_r}{\partial \theta} \right]$$

$$\nabla^2\Phi = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\Phi}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\Phi}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2\Phi}{\partial\phi^2}$$

All partials of unit vectors are zero except

$$\frac{\partial\bar{a}_r}{\partial\theta} = \bar{a}_\theta \qquad \frac{\partial\bar{a}_\theta}{\partial\theta} = -\bar{a}_r$$

$$\frac{\partial\bar{a}_r}{\partial\phi} = \bar{a}_\phi\sin\theta \qquad \frac{\partial\bar{a}_\theta}{\partial\phi} = \bar{a}_\phi\cos\theta \qquad \frac{\partial\bar{a}_\phi}{\partial\phi} = -(\bar{a}_r\sin\theta + \bar{a}_\theta\cos\theta)$$

Vector Identities. Most of the following vector identities have appeared in previous discussions. They will be useful throughout the book. Φ and ψ represent any scalar quantities, $\bar{A}$ and $\bar{B}$ any vector quantities.

$$\nabla(\Phi + \psi) = \nabla\Phi + \nabla\psi$$
$$\nabla\cdot(\bar{A} + \bar{B}) = \nabla\cdot\bar{A} + \nabla\cdot\bar{B}$$
$$\nabla\times(\bar{A} + \bar{B}) = \nabla\times\bar{A} + \nabla\times\bar{B}$$
$$\nabla(\Phi\psi) = \Phi\nabla\psi + \psi\nabla\Phi$$
$$\nabla\cdot(\psi\bar{A}) = \bar{A}\cdot\nabla\psi + \psi\nabla\cdot\bar{A}$$
$$\nabla\cdot(\bar{A}\times\bar{B}) = \bar{B}\cdot\nabla\times\bar{A} - \bar{A}\cdot\nabla\times\bar{B}$$
$$\nabla\times(\Phi\bar{A}) = \nabla\Phi\times\bar{A} + \Phi\nabla\times\bar{A}$$
$$\nabla\times(\bar{A}\times\bar{B}) = \bar{A}\nabla\cdot\bar{B} - \bar{B}\nabla\cdot\bar{A} + (\bar{B}\cdot\nabla)\bar{A} - (\bar{A}\cdot\nabla)\bar{B}$$
$$\nabla\cdot\nabla\Phi = \nabla^2\Phi$$
$$\underline{\nabla\cdot\nabla\times\bar{A} = 0}$$
$$\nabla\times\nabla\Phi = 0$$
$$\nabla\times\nabla\times\bar{A} = \nabla(\nabla\cdot\bar{A}) - \nabla^2\bar{A}$$

3

SOLUTIONS TO STATIC FIELD PROBLEMS

BASIC CONSIDERATIONS IN SOLVING FIELD PROBLEMS BY DIFFERENTIAL EQUATIONS

3.01 Introduction

Chapter 2 presented the laws of electricity and magnetism for systems with no time variations, and the concepts of such static systems. It was noted at the beginning of that chapter that it is often necessary to solve problems involving the laws of static systems, not alone for cases involving D-C potentials and direct currents but also for the many cases to be discussed later when the results of static solutions may be applied directly to the high-frequency problems of more interest to radio engineers.

If the problem is the solution of a static system, the desired result may be the actual distribution of fields or potentials, as for instance when the maximum gradient is desired for purposes of calculating breakdown voltage between a given set of electrodes. If the static solution is to be used in studying motion of electrons, it will be desired to find the field strength at a given point in space so that forces exerted on the electrons may be calculated at that point. If it is desired to use static solutions for the calculation of inductances and capacitances, or the impedance of a transmission line, it is first necessary to find the field distribution around the desired geometric configuration. Thus the first step must always be the calculation of field or potential at a given point or at all points about the electrodes, transmission line, or circuit element of interest. This will then be the goal of the present chapter. The use of the field calculations in arriving at the solutions to complete problems of various types will be considered in later chapters.

The distribution of fields may be desired in regions containing charges, in regions containing currents, or in regions free from both charges and currents. If charges are present in free space, they cannot be in equilibrium, but must be in motion; consequently this part of the problem would require a study of the motion of charges in fields. The solution for a current-carrying region is of specific interest when it is desired to calculate the impedance of a circuit element. This aspect of the problem will then be reserved until we may include the effect of frequency on

the distribution of current through the conductors. There remain the cases involving distributions in charge-free and current-free regions. All field distributions may then be obtained by a solution of the one differential equation, Laplace's equation.

Laplace's equation has universal application throughout applied physics, and the mathematics of its solution has received a great amount of attention. There are consequently many special methods available for its solution. Despite the importance of this problem to radio engineers, it would be impossible to attempt completeness here in considering all these methods. We shall study here solutions applicable to certain simple and very useful geometrical configurations, stressing those methods of solution which will best provide background for similar types of solutions in wave problems to follow.

3.02 Distribution Problems Which Involve Laplace's Equation

In the previous chapter, Laplace's equation appeared first to relate the derivatives of the electrostatic scalar potential Φ at any point in charge-free space.

$$\nabla^2 \Phi = 0 \qquad [1]$$

A solution to this equation which satisfies the boundary conditions of the specified electrode configurations and applied potentials will be an equation giving the potential as a function of the space coordinates.

In electrostatics, the potential is not the only quantity which satisfies Laplace's equation. Certain components of the electric field vector $\bar{E}$ also are distributed in space in accordance with this relation. This is easily shown by recalling a few basic relations from the previous chapter. The work integral for electric fields led to the expression

$$\nabla \times \bar{E} = 0$$

If the curl of this equation is taken (see Art. 2.38),

$$\nabla \times \nabla \times \bar{E} = 0$$

or

$$\nabla(\nabla \cdot \bar{E}) - \nabla^2 \bar{E} = 0$$

For a charge-free region Eq. 2.11(8) becomes

$$\nabla \cdot \bar{E} = 0$$

so that

$$\nabla^2 \bar{E} = 0 \qquad [2]$$

The last expression is a vector equation which in general involves derivatives of unit vectors and may not be simple in form. (See Art. 2.37.)

However, in rectangular coordinates,

$$\nabla^2\bar{E} = \bar{a}_x \nabla^2 E_x + \bar{a}_y \nabla^2 E_y + \bar{a}_z \nabla^2 E_z \qquad [3]$$

so that

$$\nabla^2 E_x = 0, \quad \nabla^2 E_y = 0, \quad \nabla^2 E_z = 0 \qquad [4]$$

Thus for a charge-free region, each of the three components of electric field intensity in rectangular coordinates satisfies Laplace's equation. Expansion of $\nabla^2\bar{E}$ in cylindrical coordinates shows that the axial component of $\bar{E}$ also satisfies Laplace's equation. That is, in cylindrical coordinates

$$\nabla^2 E_z = 0 \qquad [5]$$

(but this is not true of E_r and E_ϕ). It may often be more convenient to use these components of field directly in Laplace's equation than to use the potential, as will be illustrated later by example.

TABLE 3.02

APPLICATION OF LAPLACE'S EQUATION

Condition	Quantity	Rectangular Coordinates	Cylindrical Coordinates	Spherical Coordinates
Charge-free region (static case)	Electrostatic scalar potential	Φ	Φ	Φ
	Electric field intensity	E_x, E_y, E_z	E_z	
Current-free region (static case)	Vector magnetic potential	A_x, A_y, A_z	A_z	
	Magnetic field intensity	H_x, H_y, H_z	H_z	
	Magnetic scalar potential	Φ_m	Φ_m	Φ_m
Static currents	Current density	i_x, i_y, i_z	i_z	

Similarly for magnetic fields *in a current-free region*, the curl of magnetic intensity is zero and a like derivation applies, leading to an equation similar to (2) for H. Or, by referring to Eq. 2.25(9) with $\bar{i} = 0$, the same equation is given for $\bar{A}$.

$$\nabla^2\bar{H} = 0 \quad \nabla^2\bar{A} = 0 \qquad [6]$$

It was also pointed out in Art. 2.28 that a scalar potential could be employed for magnetic fields in a current-free region; this potential satisfies Laplace's equation directly.

Finally, the problem of D-C distributions in conductors is subject to solution by this omnipresent equation. This is so since current density $\bar{i}$

is proportional to electric field intensity by Ohm's law

$$\bar{\imath} = \sigma \bar{E} \tag{7}$$

Since the curl of $\bar{E}$ is zero, so is the curl of $\bar{\imath}$; the divergence of current must also be zero under static conditions, so by steps similar to those used for $\bar{E}$ above,

$$\nabla^2 \bar{\imath} = 0 \tag{8}$$

The above applications are summarized in Table 3.02.

3.03 Uniqueness of a Solution

Many possible means of obtaining solutions to Laplace's equation will be presented in following sections. It is important to realize that when a solution to the equation within a region is obtained, it is the only possible solution if it satisfies the boundary conditions about that region. To show that this is so, imagine that we are wrong and that there are indeed two such possible solutions, Φ_1 and Φ_2. Since they must both reduce to the same potential along the boundary,

$$\Phi_1 - \Phi_2 = 0 \tag{1}$$

along the boundary surface. Since they are both solutions to Laplace's equation,

$$\nabla^2 \Phi_1 = 0 \quad \text{and} \quad \nabla^2 \Phi_2 = 0$$

or

$$\nabla^2(\Phi_1 - \Phi_2) = 0 \tag{2}$$

throughout the entire region.

In the divergence theorem, Eq. 2.12(3), $\bar{F}$ may be any vector quantity. In particular, let it be the quantity

$$(\Phi_1 - \Phi_2)\nabla(\Phi_1 - \Phi_2)$$

Then

$$\int_V \nabla \cdot [(\Phi_1 - \Phi_2)\nabla(\Phi_1 - \Phi_2)]\, dV = \int_S [(\Phi_1 - \Phi_2)\nabla(\Phi_1 - \Phi_2)] \cdot \overline{dS}$$

From the vector identity (Art. 2.38),

$$\text{div}\,(\psi\bar{A}) = \psi \,\text{div}\, \bar{A} + \bar{A} \cdot \text{grad}\, \psi$$

the equation may be expanded to

$$\int_V (\Phi_1 - \Phi_2)\nabla^2(\Phi_1 - \Phi_2)\, dV + \int_V [\nabla(\Phi_1 - \Phi_2)]^2\, dV$$
$$= \int_S (\Phi_1 - \Phi_2)\nabla(\Phi_1 - \Phi_2) \cdot \overline{dS}$$

The first integral must be zero by (2); the last integral must be zero, since (1) holds over the boundary surface. There remains

$$\int_V [\nabla(\Phi_1 - \Phi_2)]^2 \, dV = 0 \qquad [3]$$

The gradient of a scalar is a real quantity. Thus its square can only be positive or zero. If its integral is to be zero, it can only be zero.

$$\nabla(\Phi_1 - \Phi_2) = 0 \qquad [4]$$

or

$$(\Phi_1 - \Phi_2) = \text{Constant} \qquad [5]$$

This constant must apply even to the boundary, where we know that (1) is true. The constant is then zero, and $\Phi_1 - \Phi_2$ is everywhere zero, which means that Φ_1 and Φ_2 are identical potential distributions. Hence the proof of uniqueness: Laplace's equation can have only one solution which satisfies the boundary conditions of the given region. If by any sort of conniving we find a solution to a field problem that fits all specific conditions and Laplace's equation, we may be sure it is the only one.

Problem 3.03. Prove that if charge density ρ is given throughout a volume, any solution of Poisson's equation, 2.17(3), must be the only possible solution provided that it satisfies the boundary conditions around the region.

3.04 Simple Example: Field between Coaxial Cylinders

As the first step in the study of problems in which Laplace's equation is used to obtain field distributions, it will first be applied to a very simple example. The student may check the following result by taking advantage of the symmetry of the problem and obtaining fields directly, exactly as in Art. 2.30, thus gaining confidence in the use of the equation before it is applied to more difficult geometrical configurations.

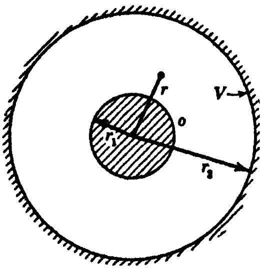


FIG. 3.04. Two coaxial conducting cylinders.

The problem is that of obtaining the field distribution between two coaxial conducting cylinders of circular cross sections with the inner cylinder at zero potential, and the outer at potential V. The geometrical symmetry (Fig. 3.04) indicates that

$$\nabla^2 \Phi = 0$$

should be expressed in cylindrical coordinates (Art. 2.38)

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial \Phi}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 \Phi}{\partial \phi^2} + \frac{\partial^2 \Phi}{\partial z^2} = 0 \qquad [1]$$

The uniformity along the axis and with respect to azimuthal angle means, of course, that nothing changes as the two corresponding coordinates are changed and so the derivatives with respect to z and ϕ are eliminated, leaving

$$\frac{\partial}{\partial r}\left(r\frac{\partial \Phi}{\partial r}\right) = 0 \qquad [2]$$

or

$$r\frac{\partial \Phi}{\partial r} = C_1 \qquad [3]$$

Integrating again

$$\Phi = C_1 \ln r + C_2 \qquad [4]$$

The constants C_1 and C_2 are still quite arbitrary, so that (4) represents any number of solutions to Laplace's equation, one for each of the possible combinations of values for C_1 and C_2. Conditions on the boundary surfaces must determine the proper solution for this case. In the present problem, these conditions are:

$$\Phi = 0 \quad \text{at} \quad r = r_1$$

$$\Phi = V \quad \text{at} \quad r = r_2$$

Thus

$$0 = C_1 \ln r_1 + C_2$$

$$V = C_1 \ln r_2 + C_2$$

Solving,

$$C_1 = \frac{V}{\ln\left(\frac{r_2}{r_1}\right)} \qquad C_2 = -\frac{V \ln r_1}{\ln\left(\frac{r_2}{r_1}\right)} \qquad [5]$$

And the potential distribution becomes

$$\Phi = \frac{V \ln\left(\frac{r}{r_1}\right)}{\ln\left(\frac{r_2}{r_1}\right)} \qquad [6]$$

Problem 3.04. Solve Laplace's equation to give the potential distribution between two concentric spheres, the inner at potential zero, the outer at potential V.

TECHNIQUES SUITABLE FOR TWO-DIMENSIONAL PROBLEMS

3.05 Graphical Field Mapping Methods

For certain configurations it is not possible or at least not desirable to obtain a mathematical solution exactly satisfying boundary conditions. A very effective graphical method exists which offers a means for obtaining approximate field distributions in a relatively short time. The method is particularly chosen for presentation here because it offers one of the best aids toward visualizing the distribution problem. It is also a very useful engineering tool, since the configurations of electrodes in vacuum tubes, electron guns, magnetic deflecting coils, etc., chosen by certain practical or mechanical considerations, are often not simple mathematical surfaces.

The method is based upon properties of the fields with which we are already familiar. It has been shown that the equipotentials and electric

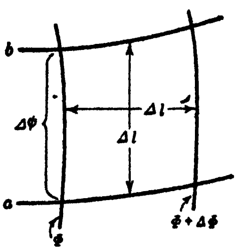


FIG. 3.05a. Element of a field map.

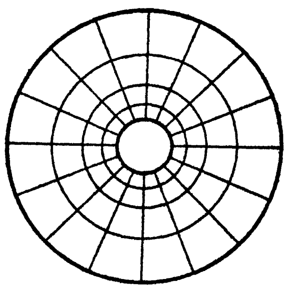


FIG. 3.05b. Map of fields between coaxial conducting cylinders.

field vectors always intersect at right angles, and the concept of flux tubes or flux lines having the direction of the electric field vector has been explained. Thus for a two-dimensional case, that is, one in which there are no variations in one direction so that the field configuration over any plane perpendicular to this direction is the same as that over any other such plane, the intersections of the equipotential lines and flux lines must divide the plane into curvilinear rectangles. Since the amount of flux each tube is said to represent (Art. 2.08) is arbitrary, it is convenient to choose the spacing between lines such that the curvilinear rectangles will reduce to curvilinear squares. The amount of flux represented by each such tube may then be determined as follows (Fig. 3.05a). If the interval between succeeding potential lines Δl apart represents a

difference of potential $\Delta\Phi$ the electric field vector or gradient has a magnitude

$$\frac{\Delta\Phi}{\Delta l}$$

The displacement vector, normal to the equipotential, then has magnitude

$$D = \frac{\epsilon'\Delta\Phi}{\Delta l}$$

The flux $\Delta\psi$ flowing through the tube bounded by the two lines a and b is

$$\Delta\psi = \bar{D} \cdot d\bar{S} = \frac{\epsilon'\Delta\Phi}{\Delta l}\,(\Delta l) = \epsilon'\Delta\Phi \qquad [1]$$

All these values are for a unit length perpendicular to the plane considered.

The procedure in making such a plot is to divide the known potential difference between electrodes into a certain number of equal intervals and to start the small curvilinear squares plot in a region where the field may be uniform, or of some form fairly accurately guessed. The plot thus proceeds throughout the region, subject to the boundary conditions that all conducting surfaces are equipotentials, and that all flux lines must enter these conducting surfaces perpendicularly. In first attempts to map the region, the condition of orthogonality of flux lines and equipotentials should always be placed first. These preliminary attempts will then not result in squares, but in curvilinear rectangles in many places. The shape of these will indicate how the plot must be revised to possess only curvilinear squares. The plot should become quite accurate after two or three such revisions. Moderately small divisions should be selected for accuracy, and any regions of low field intensity may be further subdivided without subdividing the entire plot.

Before attempting to apply this very useful method, it will be well for the student to study plots of actual configurations, and to attempt plots of simple regions where the field is known. More detailed instructions are given in several of the references, with examples of plots. Figure 3.05*b* gives the plot that would be obtained for the coaxial cylinders already solved mathematically in Art. 3.04, and Fig. 3.05*c* gives a plot made as a step in determining the characteristic impedance of a transmission line whose conductors were so shaped as to make difficult an exact mathematical solution. Note from the last example that considerations of symmetry may make unnecessary a complete plot; a plot of only a quarter of the region gives the complete information.

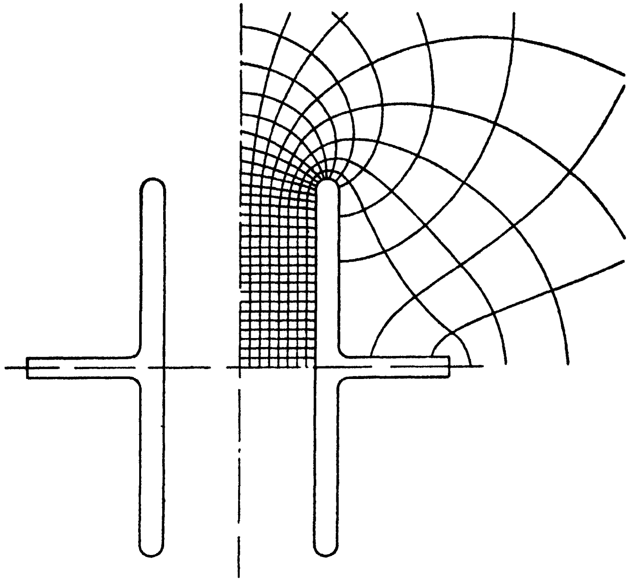


FIG. 3.05c. Map of fields between transmission line conductors of special shape.

Problem 3.05(a). Map fields between an infinite flat plane conductor and a second stepped conductor at potential V with respect to the first. The stepped conductor is a parallel plane at height a above the plane for $z < 0$, and height b for $z > 0$. The step occurs at $z = 0$. (z is measured parallel to the planes.) Take $a/b = \frac{3}{4}, \frac{1}{2}, \frac{1}{4}$.

Problem 3.05(b). Map fields between an infinite flat plane and a cylindrical conductor parallel to the plane. The conductor has diameter d and its axis is at height h above the plane. Take $d/h = 1, \frac{1}{4}$.

3.06 Introduction to Solution by Complex Function Theory

A very general mathematical attack for the two-dimensional field distribution problem utilizes the theory of functions of a complex variable. A complete treatment cannot be given here, but the method is applicable to such a variety of problems that enough of the general approach will be given so that the value of the method can be properly appreciated. It should be possible from this and the following articles to become familiar enough with the theory to study more complete treatments in the references if necessary.

In the theory of complex variables, the complex notation introduced in Chapter 1 is retained, the imaginary number $\sqrt{-1}$ being denoted by j. Thus any pure imaginary $\sqrt{-b^2}$ may be written as jb, where b is a pure real. The sum of a pure real and a pure imaginary, as $a + jb$, is called

a complex number. The variable $Z = x + jy$, where both x and y are real variables, is known as a *complex* variable. Now it happens that

$$W = u + jv$$

where u and v are each real functions of x and y, can often be neatly expressed as a function of Z, just as any real variable y may often be expressed as a function of some other real variable x.

$$W = f(Z) \tag{1}$$

There are many different complex functions, but the most interesting to us are those classed as analytic. These are the most useful since they are defined as those functions which make the derivative dW/dZ unique, where

$$\frac{dW}{dZ} = \lim_{\Delta Z \to 0} \frac{\Delta W}{\Delta Z} = \lim_{\Delta Z \to 0} \frac{f(Z + \Delta Z) - f(Z)}{\Delta Z} \tag{2}$$

Analytic functions do not include all functions since

$$\Delta Z = \Delta x + j\Delta y$$

has two degrees of freedom and hence may approach zero along any one of several paths. It is required that the value of the derivative obtained be independent of this path if the derivative is to be unique and the function analytic. Among the many properties which result from this definition, the most interesting to us will be those embodied in two differential equations which will now be derived.

$$\begin{aligned} \frac{\partial W}{\partial x} &= \frac{\partial W}{\partial Z} \cdot \frac{\partial Z}{\partial x} \\ &= \frac{dW}{dZ}\frac{\partial Z}{\partial x} \end{aligned} \tag{3}$$

since W is a function of Z only. But

$$\frac{\partial W}{\partial x} = \frac{\partial u}{\partial x} + j\frac{\partial v}{\partial x}; \quad \frac{\partial Z}{\partial x} = 1$$

so

$$\frac{dW}{dZ} = \frac{\partial u}{\partial x} + j\frac{\partial v}{\partial x} \tag{4}$$

Similarly,

$$\frac{\partial W}{\partial y} = \frac{\partial u}{\partial y} + j\frac{\partial v}{\partial y} = \frac{dW}{dZ}\frac{\partial Z}{\partial y} = j\frac{dW}{dZ}$$

If these two values of dW/dZ are to be equal, reals and imaginaries may be equated separately, requiring that

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \tag{5}$$

$$\frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} \tag{6}$$

It can be shown that these two equations, known as the Cauchy-Riemann conditions, are the necessary conditions that

$$u + jv = f(x + jy)$$

be analytic.

Differentiate (5) with respect to x, (6) with respect to y, and add

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \tag{7}$$

This is Laplace's equation in u for two dimensions. Similarly, reversing the differentiations,

$$\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} = 0 \tag{8}$$

This is Laplace's equation in v for two dimensions. Thus, since the real and imaginary parts of an analytic function of a complex variable must each satisfy Laplace's equation, we are in possession of a huge store of solutions to the equation in the two-dimensional case. For example, take the function

$$W = AZ + B$$

or

$$u + jv = A(x + jy) + B$$

where A and B are reals. If real and imaginary parts are separated,

$$u = Ax + B$$

$$v = Ay$$

Both u and v are known from the foregoing theory to be solutions to Laplace's equation. Thus both u and v are possible potentials; that is, for two different particular problems, u and v will be the functions that tell how the potential is distributed. It is not difficult here to see what the two particular problems are. For

$$\Phi = u = Ax + B$$

we get a uniform field in the x direction, and for

$$\Phi = v = Ay$$

a uniform field in the y direction. In a similar (but more complicated) manner

$$W = \sin Z, \quad W = \ln Z, \quad W = Z^2$$

give rise to two solutions each to Laplace's equation and thus represent potential distributions corresponding to certain boundary conditions.

The mere possession of a tool for obtaining a large number of solutions to Laplace's equation is not sufficient; we must understand how the proper one of these solutions is to be chosen for an actual problem. For this aspect of the problem, it will be helpful to consider first the physical significance that may be attached to u and v.

3.07 Flow and Potential Functions

The Cauchy-Riemann conditions, Eqs. 3.06(5) and 3.06(6), if multiplied together, give

$$\left(\frac{\partial u}{\partial x}\right)\left(\frac{\partial u}{\partial y}\right) = -\left(\frac{\partial v}{\partial x}\right)\left(\frac{\partial v}{\partial y}\right) \tag{1}$$

These are precisely the conditions necessary to require that the curves u = constant and v = constant, plotted in the x-y plane, should intersect at right angles. That this is so is proven in most analytic geometry books in something like the following way. For the curves on which u is a constant, $du = 0$ or

$$\left[\frac{\partial u}{\partial x}\,dx + \frac{\partial u}{\partial y}\,dy\right]_{u=\text{const.}} = 0$$

the equation of a curve whose slope is obviously

$$\left.\frac{dy}{dx}\right|_{u=\text{const.}} = -\frac{\partial u/\partial x}{\partial u/\partial y}$$

Similarly,

$$\left.\frac{dy}{dx}\right|_{v=\text{const.}} = -\frac{\partial v/\partial x}{\partial v/\partial y}$$

For the two curves to intersect at right angles

$$\left.\frac{dy}{dx}\right|_{u=\text{const.}} = -\left.\frac{dx}{dy}\right|_{v=\text{const.}}$$

or

$$-\frac{\partial u/\partial x}{\partial u/\partial y} = \frac{\partial v/\partial y}{\partial v/\partial x} \tag{2}$$

By use of the Cauchy-Riemann conditions, (1) is seen to be identical with (2).

Suppose now that for some problem u is the distribution of potential, i.e., the curves u = constant are equipotential lines. Then from the orthogonality relation just determined, lines of constant v must intersect lines of constant u (equipotentials) at right angles, and may hence be properly interpreted as flux lines. The real part of any analytic function of a complex variable may then be considered as the potential function with the imaginary part as the corresponding flux or flow function,

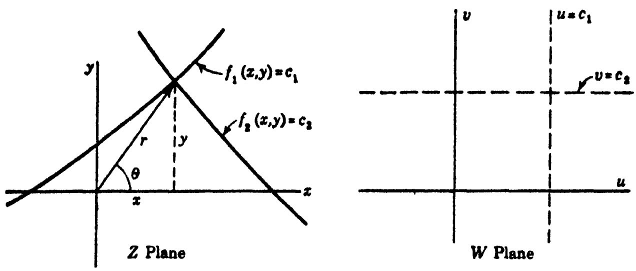


FIG. 3.07. Transformations between Z and W plane.

or the imaginary part might be taken as the potential function (for a different problem of course) with the real part as the corresponding flow function, each of these situations corresponding to some set of boundary conditions.

This may be better visualized if a complex quantity

$$Z = (x + jy)$$

is plotted as the point (x,y) on the Z plane (Fig. 3.07). The coordinates of the point may also be written in polar coordinates as (r,ϕ) where

$$r = \sqrt{x^2 + y^2}, \quad \phi = \tan^{-1}\left(\frac{y}{x}\right)$$

Strictly speaking, Z on such a plot is the radius vector from the origin to the point (x,y) or (r,ϕ) since it is correctly expressed by

$$Z = x + jy = re^{j\varphi}$$

This convention is quite familiar to electrical engineers since in the representation of voltages and current as complex quantities, these are plotted in a similar manner. To be consistent with the nomenclature used in

such cases, r may be known as the magnitude and ϕ as the phase angle of the complex quantity. Similarly,

$$W = u + jv$$

is plotted as the point (u,v) on the W plane. The lines $u =$ constant and $v =$ constant are thus two sets of orthogonal lines parallel to the u axis and v axis respectively on the W plane. But as

$$W = f(Z) = u + jv = f(x + jy)$$
$$u = f_1(x,y)$$
$$v = f_2(x,y)$$

so the lines $u =$ constant transform to curves on the Z plane given by $f_1(x,y) =$ constant. Similarly, $v =$ constant transforms to a second set of curves orthogonal to the first set on the Z plane. To summarize, for a given function

$$W = f(Z),$$

any point (u,v) on the W plane corresponds to some point (x,y) in the Z plane. The exact correspondence between the two points is obtained directly from the equation

$$W = f(Z)$$

Every curve $u = F(v)$ in the W plane corresponds to some curve

$$x = F_1(y)$$

in the Z plane; every region in the W plane corresponds to some region in the Z plane. The function

$$W = f(Z)$$

which accomplishes this transformation from corresponding points and regions from W plane to Z plane is often called simply the transformation. Figure 3.07 shows corresponding curves as described.

3.08 Identifying the Problem with the Transformation

As an example of how the curves $u =$ constant and $v =$ constant from some transformation

$$W = f(Z)$$

may be identified with the problem in field theory for which it is a solution, consider the transformation

$$W = \cos^{-1} Z$$

or

$$u + jv = \cos^{-1}(x + jy) \quad [1]$$

It should first be determined that this is an analytic function, and the Cauchy-Riemann conditions show that it is. Incidentally, it has been found from mathematical study that if W as a function of Z is any of the simple functions (algebraic, sinusoidal, hyperbolic, exponential) it is always analytic, so we may dismiss this point in following examples.

Expanding,

$$x + jy = \cos(u + jv) = \cos u \cosh v - j \sin u \sinh v$$

$$x = \cos u \cosh v$$

$$y = -\sin u \sinh v$$

It then follows that

$$\frac{x^2}{\cosh^2 v} + \frac{y^2}{\sinh^2 v} = 1 \qquad [2]$$

$$\frac{x^2}{\cos^2 u} - \frac{y^2}{\sin^2 u} = 1 \qquad [3]$$

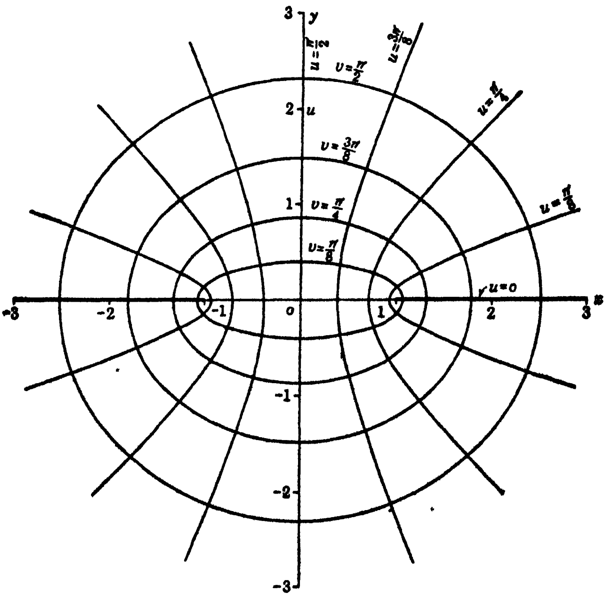


FIG. 3.08. Plot of the transformation $u + jv = \cos^{-1}(x + jy)$.

Equation (2) represents a set of confocal ellipses and (3) represents a set of confocal hyperbolas, orthogonal to the ellipses. These are plotted on Fig. 3.08. Thus the $v \doteq$ constant lines of this transformation could

be used to represent equipotentials about a conducting cylinder of ellipsoidal cross section; the u = constant lines could be used to represent potential lines between two hyperbolic cylindrical electrodes at different potentials. With a proper choice of the region, and the function (either u or v) to serve as potential along with the particular curve or curves on which the potential is to be specified, the transformation $W = \cos^{-1} Z$ gives the solution to all the following problems:

1. Field around a charged elliptic conducting cylinder.
2. Field between hyperbolic cylinders.
3. Field between two semi-infinite conducting plates, coplanar and with a slit separating them. (This case degenerates from (2) above.)
4. Field between perpendicular semi-infinite plates separated by a gap.
5. Field between a plate and a surrounding elliptic cylinder. (The plate is a degenerate ellipse.)
6. Field between a hyperbolic cylinder and a plate.

Usually an actual problem requires the reverse of the above procedure. That is, given the physical shape of electrodes with applied potentials, it is desired to find the function

$$W = f(Z)$$

which will transform the curves corresponding to conducting boundaries in the Z plane to lines parallel to one of the axes in the W plane, i.e., to u = constant or v = constant lines. This is usually not such a simple problem. It will be carried out in some special cases which follow to show some of the considerations that may lead to the desired transformation.

3.09 Examples of Transformations Set Up to Fit Geometrical Configurations

A. Intersection of Planes. Suppose that it is desired to find the field distribution in the neighborhood of the angle formed by two intersecting grounded conducting planes (Fig. 3.09a). This field presumably arises from a difference of potential between this angular conductor and some other conductor far removed. If the axes of x and y in the Z plane are selected as shown and we wish to make the lines v = constant correspond to equipotentials, the zero potential boundary AOB should transform to the line $v = 0$ in the W plane. It appears that this might be accomplished by expanding the region inside AOB as the opening of a fan, until AOB becomes a straight line. If the polar forms

$$Z = re^{j\theta}$$

and

$$W = \rho e^{j\phi}$$

are used, we can deduce that the mathematical expression

$$W = (Z)^{\pi/\alpha} = r^{\pi/\alpha} e^{j\frac{\pi\theta}{\alpha}} \qquad [1]$$

expresses this transformation, since when $\theta = \alpha$, $\phi = \pi$. Thus any point lying along the line $\theta = \alpha$ in the Z plane transforms into a point

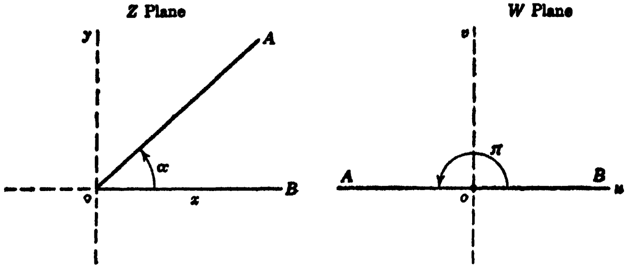


FIG. 3.09a. Transformation $W = (Z)^{\pi/\alpha}$ applied to a wedge of angle α.

along the negative u axis in the W plane. Lines corresponding to u = constant and v = constant according to the function

$$u + jv = (x + jy)^{\pi/\alpha} \qquad [2]$$

thus plot the flux lines and equipotentials respectively for the region inside the angle.

B. Parallel Cylindrical Conductors. Transformations will be deduced for a second case, Fig. 3.09e, although here the results are not obtained so directly. Consider first a line charge passing through the origin of the Z plane (Fig. 3.09b). Start by choosing u as the potential function, v as the flow function. Symmetry requires that the constant potentials (u = constant) be circles, the flow lines (v = constant) straight lines through the origin. The equation for the potential function outside the charge must be the same as that for the simple cylindrical conductor case (Art. 3.04),

$$u = C_1 \ln r$$

Thus whatever transformation is used should give the relation

$$u + jv = C\,(\ln r + j\theta) \qquad [3]$$

Then the transformation

$$W = C \ln Z \qquad [4]$$

accomplishes exactly this, since

$$\ln Z = \ln\,(re^{j\theta}) = \ln r + j\theta$$

If the line charge is not on the origin, but at a distance a from it (Fig. 3.09c), the problem is reduced to the original case by a simple change of variables, so the transformation should now differ from the first only in that Z is replaced by $Z - a$.

$$W = C \ln (Z - a) \qquad [5]$$

That this does lead to the desired configuration is shown by expanding

$$u + jv = C\left(\tfrac{1}{2} \ln [(x - a)^2 + y^2] + j \tan^{-1} \frac{y}{x - a}\right)$$

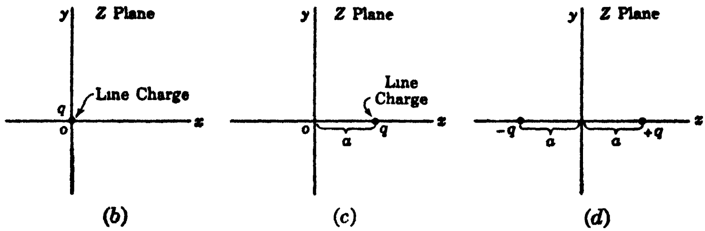


FIG. 3.09.

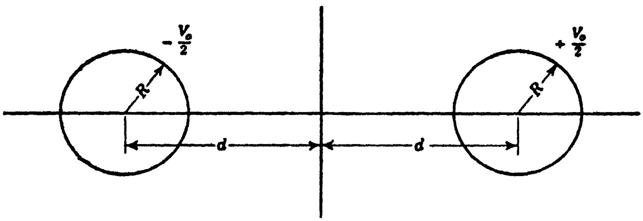


FIG. 3.09e. Two parallel conducting cylinders.

The lines u = constant are curves $(x - a)^2 + y^2$ = constant, representing a family of circles with $(a,0)$ as center. The lines v = constant become $\dfrac{y}{x - a}$ = constant, representing a family of straight lines, all passing through $(a,0)$.

If there are next two line charges of opposite sign but equal magnitudes, one at $x = +a$, and the other at $x = -a$ (Fig. 3.09d), the correct new transformation can be obtained by superposition; since superposition holds for both the potential and flow functions, the resultant potential and flow functions should be the sum of the two from the individual line charges. Thus

$$W = C[\ln (Z - a) - \ln (Z + a)] \qquad [6]$$

should be the transformation. Expanding and taking the real part,

$$u = \frac{C}{2} \ln \frac{[(x - a)^2 + y^2]}{[(x + a)^2 + y^2]} \tag{7}$$

The lines u = constant then correspond to

$$\frac{(x - a)^2 + y^2}{(x + a)^2 + y^2} = K$$

or

$$y^2 + \left[x - \frac{a(1 + K)}{1 - K}\right]^2 = \frac{a^2 4K}{(1 - K)^2} \tag{8}$$

This equation for equipotentials is that of a family of circles with center at

$$x = \frac{a(1 + K)}{1 - K}$$

and radii $2a\sqrt{K}/(1 - K)$. We might replace any of these constant potential circles by a circular conductor of appropriate potential. That is, if R is the radius of such a conductor, d the distance of its center from the origin, then the values of K and a in terms of R and d can be calculated from the relations

$$\frac{a(1 + K)}{1 - K} = d \qquad \frac{2a\sqrt{K}}{1 - K} = R \tag{9}$$

It follows that (7) can be used to express the equation for equipotentials between two parallel cylindrical conductors, as shown in Fig. 3.09*e*. The constant a for this equation is obtainable in terms of dimensions by (9); C is obtainable in terms of V_0, say by setting $u = V_0/2$ at $x = d - R$, $y = 0$. The result is

$$u = \Phi = \frac{V_0}{2} \left\{ \frac{\ln [(x - a)^2 + y^2] - \ln [(x + a)^2 + y^2]}{\ln [d - R - a]^2 - \ln [d - R + a]^2} \right\} \tag{10}$$

where

$$a = \sqrt{d^2 - R^2}$$

3.10 Transformations for Polygons in General

For the general method of obtaining two-dimensional field distributions from transformations between complex functions, two specific examples have been given in which desired transformations were obtained by a combined use of the physical picture and good guesswork.

There is a general method which may be applied to any electrode configuration having the form of a polygon with linear sides. For such cases Schwarz and Christoffel have shown that a differential equation may be written which leads directly to the desired transformation. It will not be derived here, nor will all its applications be considered thoroughly, but a brief discussion will aid in the appreciation of the transformation method generally, and also place the student in an advantageous position to read further in the references given.

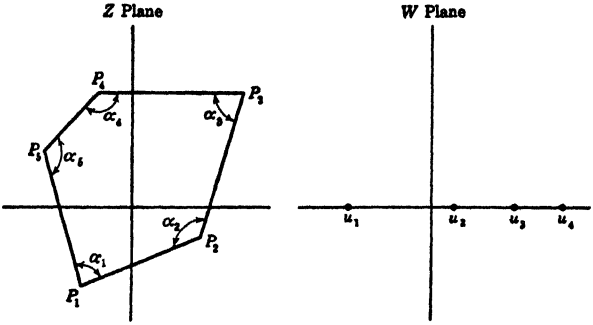


FIG. 3.10. Transformation from a general polygon to a straight line.

Consider any polygon which is the equipotential conducting boundary of a region (Fig. 3.10). The desired transformation function should transform this boundary into a line u = constant or v = constant in the W plane. If the line $v = 0$ is selected, the vertex P_1 in the Z plane will correspond to some point u_1 in the W plane, P_2 to u_2, etc. A consideration first of one vertex alone, as P_1 and its corresponding point u_1, shows that the problem is similar to the case of Art. 3.09 where the angle formed by the intersection of two planes was straightened out to the line $v = 0$ in the W plane by the function

$$W = Z^{\pi/\alpha} \quad \text{or} \quad Z = W^{\alpha/\pi}$$

Then

$$\frac{dZ}{dW} = \frac{\alpha}{\pi} W^{\frac{\alpha}{\pi}-1} \qquad [1]$$

α is the interior angle as in Art. 3.09.

In the above transformation, the vertex of the angle is transformed to the origin of the W plane. If it is to correspond instead to the point u_1, it is only necessary to perform a simple change of variables, $(W - u_1)$ for W above,

$$\frac{dZ}{dW} = \text{constant} \times (W - u_1)^{\frac{\alpha_1}{\pi}-1} \qquad [2]$$

The extension to a large number of vertices is

$$\frac{dZ}{dW} = K(W - u_1)^{\frac{\alpha_1}{\pi}-1}(W - u_2)^{\frac{\alpha_2}{\pi}-1}\cdots(W - u_n)^{\frac{\alpha_n}{\pi}-1} \qquad [3]$$

Note in particular the phase angle of dZ/dW. As one progresses along the u axis, the factors $(W - u_m)$ can have a phase of only π if $W < u_m$, or 0 if $W > u_m$. Thus the factor $(W - u_m)^{\frac{\alpha_m}{\pi}-1}$ changes phase by $(\pi - \alpha_m)$ as W passes through the point u_m in a direction of increasing W. This is the only change in phase of dZ/dW at this point, so it corresponds as required to a straightening out of the angle at P_m to a straight line at u_m. There is no further change in phase of dZ/dW until the next singular point is reached, which is as expected since this region corresponds to a straight side of the polygon.

The desired functional relation between Z and W is obtained by an integration of (3). The interior of the polygon in the Z plane then corresponds to the upper half of the W plane.

The Schwarz transformation is most often used as only one of the transformations required for a given problem. For example, two of the types of problems for which the Schwarz transformation is useful are:

(*a*) In obtaining field distributions between a conducting polygon and an infinitesimal wire on the interior, raised to a different potential from the polygon.

(*b*) In finding field distributions between two separate conductors at different potentials, these conductors extending to infinity so that they may be considered as closed at infinity. (Examples in Table 3.10 are of this type.)

In problems of type (*a*) the polygon may be transformed to the horizontal axis of an auxiliary plane, say S plane, and the wire will then correspond to some point in the upper half of this plane. The problem in the S plane is then that of an infinitesimal wire parallel to a plane and may be solved readily (say by the method of images to follow in Art. 3.80). This solution for the S plane then corresponds to the transformation from the S to the W plane, and all variables of the S plane may be eliminated, leaving W as a function of Z as required.

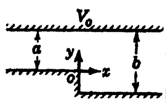
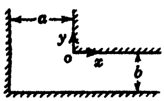
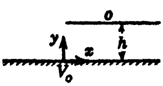
In problems of type (*b*), it is most often convenient to transform one of the equipotential boundaries to the left half of the horizontal axis in the S plane, the other to the right half; since the solution for field distributions between two semi-infinite planes separated by an infinitesimal gap is readily obtained, the problem in the S plane is again easily solved.

Some examples of typical problems with their transformations are

listed in Table 3.10. In these $W = u + jv$ where u is the flux function, v the potential function, and thus potential and flux distributions as functions of the coordinates x and y are obtainable by separating real and imaginary parts of the transformations.

TABLE 3.10

$Z = x + jy$; $W = u + jv$ WHERE u = FLUX FUNCTION, v = POTENTIAL

	$Z = \frac{b}{\pi}\left\{\cosh^{-1}\left[\frac{\alpha^2+1-2\alpha^2e^{kW}}{1-\alpha^2}\right] - \alpha\cosh^{-1}\left[\frac{2e^{-kW}-(\alpha^2+1)}{1-\alpha^2}\right]\right\}$ $\alpha = \frac{a}{b}$ $\quad k = \frac{\pi}{V_0}$
	$Z = \frac{b}{\pi}\left[\ln\left(\frac{1+S}{1-S}\right) - 2\alpha\tan^{-1}\left(\frac{S}{\alpha}\right)\right]$ $\alpha = \frac{a}{b}$ $\quad S = \sqrt{\frac{e^{kW}+\alpha}{e^{kW}-1}}$
	$Z = \frac{h}{\pi}[e^{kW} - kW + j\pi]$ $k = \frac{\pi}{V_0}$

TECHNIQUES FOR SOLVING THREE- (OR TWO-) DIMENSIONAL PROBLEMS

3.11 The Product Solution Method for Three-Dimensional Problems

Most of the space given to the two-dimensional problems has been used for discussion of the transformation method. The three-dimensional case has no equivalent. However, there is an attack which is generally useful for three-dimensional problems — one which is possibly even easier to visualize than the two-dimensional transformation method. In this method, it is found that solutions may be obtained to Laplace's equation for three dimensions (in either rectangular, cylindrical, or spherical coordinates) of a form such that variables are sepa-

rated. The procedure is the common one for solving a partial differential equation by assuming that its solution may be expressed as a product of functions, each containing only one of the variables of the coordinate system used. (For example, in a cylindrical coordinate system, solution for potential may be expressed as a product of three functions, one of the radial distance r, one of the azimuthal angle ϕ, and one of the axial distance z.) Substitution in the partial differential equation and separation of variables leads to ordinary differential equations which may be solved separately.

It may seem that elimination at the outset of all solutions not of the product form represents a severe limitation, but this is not so since we may include a sum or series of these solutions when one alone will not permit matching of the boundary conditions. This is permissible since the sum of separate solutions to a linear differential equation is also a solution of the equation. The amounts of the individual solutions to be added are determined by the boundary conditions in a manner analogous to that used to determine the amounts of the individual harmonics to add up to a complex wave shape in a Fourier analysis. (See Arts. 1.14 and 1.23.)

In following sections we shall then show how to obtain the space harmonic solutions to Laplace's equation, and how to combine these to fit, exactly or approximately, the boundary conditions for many shapes of electrodes. Analogy to such a series method will be found in many later chapters in the series wave solutions required for solution of the high-frequency field distribution problems.

3.12 Cylindrical Harmonics

A large class of problems of major interest is that in which field distribution is desired throughout a region having cylindrical symmetry about an axis. For example, there are the familiar electrostatic electron lenses found in many cathode-ray tubes, consisting of lengths of coaxial circular cylinders end to end. One of these is indicated in Fig. 3.12.

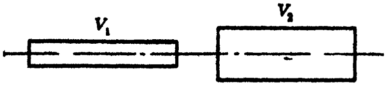


FIG. 3.12. Cross section of cylindrically symmetric electron lens.

The cylindrical shapes indicate the use of cylindrical coordinates; symmetry eliminates variations with ϕ. Thus Laplace's equation becomes (Art. 2.38)

$$\frac{\partial^2\Phi}{\partial r^2} + \frac{1}{r}\frac{\partial\Phi}{\partial r} + \frac{\partial^2\Phi}{\partial z^2} = 0 \qquad [1]$$

To solve this equation let us try to find solutions of the product form. That is, try

$$\Phi = RZ \tag{2}$$

where R is a function of r alone, Z of z alone. Substitute in the differential equation (1)

$$R''Z + \frac{1}{r}R'Z + RZ'' = 0$$

R'' denotes d^2R/dr^2, Z'' denotes d^2Z/dz^2, etc. Separate variables by dividing by RZ.

$$\frac{Z''}{Z} = -\left[\frac{R''}{R} + \frac{1}{r}\frac{R'}{R}\right]$$

This equation must be true for all values of r and z. The right side of the equation does not contain z, so it cannot vary with z; the left side cannot then vary with z either, because the right and left sides must always be equal. Similarly, the left side does not contain r, so the right side cannot. Both sides must then be a constant, the same constant. Let this constant be T^2. Two ordinary differential equations then result as follows.

$$\frac{1}{R}\frac{d^2R}{dr^2} + \frac{1}{rR}\frac{dR}{dr} = -T^2 \tag{3}$$

$$\frac{1}{Z}\frac{d^2Z}{dz^2} = T^2 \tag{4}$$

The second equation is the familiar differential equation of simple harmonic motion studied in Chapter 1. The solution is then in sinusoids if T^2 is negative, in hyperbolic functions (or exponentials) if T^2 is positive.

(*a*) First consider (3) with T^2 positive so that the solution to (4) is in terms of hyperbolic functions. Equation (3) is then

$$\frac{d^2R}{dr^2} + \frac{1}{r}\frac{dR}{dr} + T^2R = 0 \tag{5}$$

In Chapter 1, the familiar equation resulting in sinusoids [as (4) above] was solved by assuming a solution in the form of a power series. Substitution in the differential equation told the form this series must have to be truly a solution of the equation. Similarly, to solve (5), the function R may also be assumed to be some series of powers of r.

$$R = a_0 + a_1r + a_2r^2 + a_3r^3 + \cdots$$

or

$$R = \sum_{p=0}^{\infty} a_p r^p \tag{6}$$

Substitution of this function in (5) shows that it is a solution if the constants are as follows.

$$a_p = a_{2m} = C_1(-1)^m \frac{\left(\frac{T}{2}\right)^{2m}}{(m!)^2}$$

(C_1 is any arbitrary constant.) That is,

$$R = C_1 \sum_{m=0}^{\infty} \frac{(-1)^m \left(\frac{Tr}{2}\right)^{2m}}{(m!)^2} = C_1\left[1 - \left(\frac{Tr}{2}\right)^2 + \frac{\left(\frac{Tr}{2}\right)^4}{(2!)^2} - \cdots\right] \tag{7}$$

is a solution to the differential equation (5).

The series is not recognized as the series for a simple function, as were the series for sines and cosines in Chapter 1, but it is easy to check and find that it is convergent, so that values may be calculated for any argument (Tr). Such calculations have been made over a wide range of values for the argument, the results tabulated, and the function defined by the series denoted by $J_0(Tr)$ and called a Bessel function (of first kind, zero order; the reason for such specific designation will be apparent later). Thus defined,

$$J_0(v) \equiv 1 - \left(\frac{v}{2}\right)^2 + \frac{\left(\frac{v}{2}\right)^4}{(2!)^2} - \cdots \equiv \sum_{m=0}^{\infty} \frac{(-1)^m \left(\frac{v}{2}\right)^{2m}}{(m!)^2} \tag{8}$$

The particular solution (7) may then be written simply as

$$R = C_1 J_0(Tr)$$

The differential equation (5) is of the second order and so must have a second solution with a second arbitrary constant. (The sine and cosine constitute the two solutions for the simple harmonic motion equation.) By properly manipulating the series again, a second independent solution can be obtained. This may be called a Bessel function of second kind, order zero, and can have one of several forms. A form easily found

in tables is

$$N_0(v) = \frac{2}{\pi}\ln\left(\frac{\gamma v}{2}\right)J_0(v) - \frac{2}{\pi}\sum_{m=1}^{\infty}\frac{(-1)^m\left(\frac{v}{2}\right)^{2m}}{(m!)^2}\left[1+\frac{1}{2}+\frac{1}{3}+\cdots\frac{1}{m}\right] \quad [9]$$

The constant $\ln\gamma = 0.5772\cdots$ is Euler's constant. In general, then,

$$R = C_1J_0(Tr) + C_2N_0(Tr) \quad [10]$$

is the solution to (5), with

$$Z = C_3\sinh(Tz) + C_4\cosh(Tz) \quad [11]$$

as the corresponding form for the solution to (4). It should be noted from (9) that $N_0(Tr)$, the second solution to R, becomes infinite at $r = 0$, so it cannot be present in any problem for which $r = 0$ is included in the region over which the solution applies.

(*b*) If T^2 is negative, let $T^2 = -\tau^2$ or $T = j\tau$, where τ is real, and (5) may be written

$$\frac{d^2R}{dr^2} + \frac{1}{r}\frac{dR}{dr} - \tau^2R = 0 \quad [12]$$

The series (7) is still a solution, and T in (7) may be replaced by $j\tau$. Since all powers of the series are even, imaginaries disappear, and a new series is obtained which is real and also convergent. That is,

$$J_0(jv) = 1 + \left(\frac{v}{2}\right)^2 + \frac{\left(\frac{v}{2}\right)^4}{(2!)^2} + \frac{\left(\frac{v}{2}\right)^6}{(3!)^2} + \cdots \quad [13]$$

Values of $J_0(jv)$ may be calculated for various values of v from such a series; these are also tabulated in the references. The defined function is denoted $I_0(v)$ in many of the references. Thus a solution to (12) is

$$R = C_1'J_0(j\tau r) \equiv C_1'I_0(\tau r) \quad [14]$$

There must also be a second solution in this case, and since it is usually not taken simply as $N_0(j\tau r)$, the choice of this will be discussed in a later article (3.17). This second solution does become infinite at $r = 0$, however, just as does $N_0(Tr)$, and so will not be required in the simple examples immediately following which include the origin, $r = 0$, in the solution.

The solution to the z equation (4) when $T^2 = -\tau^2$ is

$$Z = C_3' \sin(\tau z) + C_4' \cos(\tau z) \qquad [15]$$

Solutions of the type (2) with R and Z given by either (10) and (11) or (14) and (15) are the cylindrical harmonics sought. They will now be applied to several simple examples.

3.13 Simple Example of a Field Described by One Cylindrical Harmonic

The form of electrode configuration for which a solution is given by only one term of any of the product forms obtained above may be determined. For example, a special case of the solution to Laplace's equation defined by Eqs. 3.12(14) and 3.12(15) is

$$\Phi = CI_0(\tau r) \sin(\tau z) \qquad [1]$$

where $I_0(\tau r) \equiv J_0(j\tau r)$. It is seen that $\Phi = 0$ for all values of r, both at $\tau z = 0$ and $\tau z = \pi$, so for this solution the region may be terminated by two conducting planes of zero potential at $z = 0$ and $z = l = \pi/\tau$. Also, since C is yet arbitrary, it may be fixed by specifying the potential at any point, say $\Phi = V_0$, at $r = r_0$, $z = l/2$, then, since

$$\tau = \frac{\pi}{l}$$

$$V_0 = CI_0\left(\frac{\pi r_0}{l}\right) \sin\left(\frac{\pi}{2}\right) = CI_0\left(\frac{\pi r_0}{l}\right)$$

or

$$C = \frac{V_0}{I_0\left(\frac{\pi r_0}{l}\right)}$$

The arbitrary constant is thus fixed. A curve $r = f(z)$ may be obtained showing the complete constant potential surface, $\Phi = V_0$, by setting

$$V_0 = CI_0(\tau r) \sin(\tau z) \qquad [2]$$

or

$$I_0\left(\frac{\pi r}{l}\right) \sin\left(\frac{\pi z}{l}\right) = I_0\left(\frac{\pi r_0}{l}\right)$$

The curve satisfying (2) is plotted in Fig. 3.13 from the tabulated values of $I_0(v)$. If a conducting boundary is placed along this curve and maintained at potential V_0, it is then evident that (1) is a proper solution for this region since all boundary conditions are satisfied. That is,

inside the region,

$$\Phi = V_0 \frac{I_0\left(\dfrac{\pi r}{l}\right)}{I_0\left(\dfrac{\pi r_0}{l}\right)} \sin\left(\frac{\pi z}{l}\right) \qquad [3]$$

Note in particular that the potential along the axis, which is often of major interest, has a simple sinusoidal distribution with z.

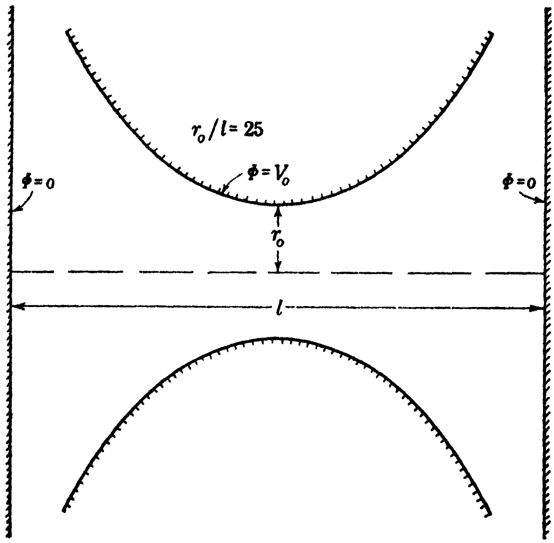


FIG. 3.13. Shape of electrodes for field described by $\Phi = CI_0\left(\frac{\pi r}{l}\right) \sin\left(\frac{\pi z}{l}\right)$.

In general, the electrode shapes of an actual physical problem may be quite different from the boundary of this example; it is then that other terms must be added to build up a potential distribution that will fit the actual boundaries. Such examples will follow.

Problem 3.13. Find the electrode configuration for which the single cylindrical harmonic

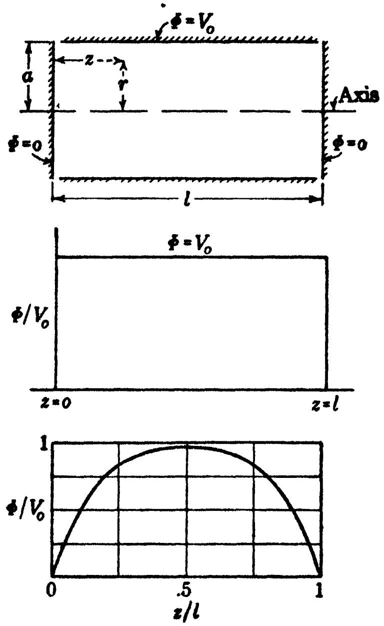
$$\Phi = CJ_0(Tr) \sinh (Tz)$$

gives the solution for potential distribution.

3.14 Example of a Field Described by a Series of Cylindrical Harmonics

If it is desired to find potential distribution inside a region bounded by a circular cylinder at potential V_0 and two planes perpendicular to the

axis at zero potential (Fig. 3.14*a*), cylindrical symmetry exists and product solutions must be obtainable from Art. 3.12. The necessary gaps between electrodes are assumed negligibly small compared to all other dimensions.



(*a*) Cylindrical region ($a/l = 0.25$).

(*b*) Plot of potential vs. z at $r = a$.

(*c*) Plot of potential vs. z at $r = 0$.

FIG. 3.14.

Potential must be zero at $z = 0$ and $z = l$ for all values of r, which helps us to select the proper form of the product solution as follows:

1. Sinusoidal solutions for z are desired rather than hyperbolic, since the latter do not have repeated zeros and such a characteristic of the solution is necessary.

2. Coefficient of the cosine term must be zero since Φ is to be zero at $z = 0$.

3. Periodicity, τ, is given by $m\pi/l$, where m may be any integer from zero to infinity, if Φ is to be zero at $z = l$.

So

$$Z = C_3 \sin \tau z = C_3 \sin\left(\frac{m\pi z}{l}\right)$$

The corresponding solution for R is Eq. 3.12(14).

$$R = C_1' I_0(\tau r) = C_1' I_0\left(\frac{m\pi r}{l}\right)$$

Thus all solutions to Laplace's equation of the product form which satisfy the symmetry of the problem and the boundary conditions so far imposed must be of the form

$$\Phi_m = A_m I_0\left(\frac{m\pi r}{l}\right)\sin\left(\frac{m\pi z}{l}\right)$$

and a series of these harmonics, having amplitudes A_m yet to be determined, will give the potential distribution Φ desired.

$$\Phi = A_0 + A_1 I_0\left(\frac{\pi r}{l}\right)\sin\left(\frac{\pi z}{l}\right) + A_2 I_0\left(\frac{2\pi r}{l}\right)\sin\left(\frac{2\pi z}{l}\right) + \cdots$$

or

$$\Phi = \sum_{m=0}^{\infty} A_m I_0\left(\frac{m\pi r}{l}\right)\sin\left(\frac{m\pi z}{l}\right) \qquad [1]$$

An additional boundary condition remains that at $r = a$, Φ is zero at $z = 0$ and $z = l$, but equal to V_0 for all other values of z. A plot of this distribution of Φ against z at $r = a$ (Fig. 3.14*b*) results in a square wave shape such as was expanded in a Fourier series in Chapter 1. From Eq. 1.23(10), this function is representable over the range $0 < z < l$ by the series

$$\Phi\bigg|_{r=a} = \frac{4V_0}{\pi}\sum_{p=1}^{\infty}\frac{1}{2p-1}\sin\left[\frac{(2p-1)\pi z}{l}\right]$$

But (1) gives $\Phi|_{r=a}$ as

$$\Phi\bigg|_{r=a} = \sum_{m=1}^{\infty} A_m I_0\left(\frac{m\pi a}{l}\right)\sin\left(\frac{m\pi z}{l}\right)$$

Since these values of $\Phi|_{r=a}$ must be equal to each other for all values of z, they must be equal term by term. That is, corresponding coefficients of $\sin(m\pi z/l)$ may be equated.

$$m = 2p - 1$$

and

$$A_{(2p-1)} I_0\left[\frac{(2p-1)\pi a}{l}\right] = \frac{4V_0}{\pi(2p-1)}$$

or

$$A_{(2p-1)} = \frac{4V_0}{\pi(2p-1) I_0\left[\dfrac{(2p-1)\pi a}{l}\right]}$$

Thus every coefficient is determined and the potential at any point in the region is given by substituting these determined coefficients in the series (1)

$$\Phi = \frac{4V_0}{\pi}\left[\frac{I_0\left(\frac{\pi r}{l}\right)\sin\left(\frac{\pi z}{l}\right)}{I_0\left(\frac{\pi a}{l}\right)} + \frac{I_0\left(\frac{3\pi r}{l}\right)\sin\left(\frac{3\pi z}{l}\right)}{3I_0\left(\frac{3\pi a}{l}\right)} + \cdots\right]$$

or

$$\Phi = \sum_{p=1}^{\infty} \frac{4V_0}{\pi(2p-1)} \frac{I_0\left[\frac{(2p-1)\pi r}{l}\right]}{I_0\left[\frac{(2p-1)\pi a}{l}\right]} \sin\left[\frac{(2p-1)\pi z}{l}\right] \qquad [2]$$

A plot of potential distribution along the axis, Φ/V_0 versus z at $r = 0$, is given in Fig. 3.14c for a case with $a/l = 0.25$.

3.15 Bessel Functions of Zero Order: Real Arguments

In the solution of Laplace's equation in cylindrical coordinates, a differential equation appeared of the form

$$\frac{d^2R}{dr^2} + \frac{1}{r}\frac{dR}{dr} + T^2R = 0 \qquad [1]$$

This equation, known as Bessel's equation, is common throughout applied physics, and in particular arises in many field problems involving cylindrical and spherical configurations. Since a large number of the structures of radio engineering, for example, vacuum tube electrodes, circular wave guides, and ordinary round wires, have such forms, the equation and its solutions will occur frequently throughout the book. Although we have already made use of its solutions in the preceding examples, we shall now devote some time to a special study of the properties of its solutions. These solutions to Bessel's equation are called Bessel functions.

Equation (1) is quite similar to the equation of simple harmonic motion. This familiar differential equation,

$$\frac{d^2Z}{dz^2} + K^2Z = 0 \qquad [2]$$

was studied in Chapter 1 and found to have solutions in sines and cosines.

$$Z = A\cos Kz + B\sin Kz \qquad [3]$$

The solutions to the Bessel equation, (1), were obtained in Art. 3.12 by a

method also used for solving the simple harmonic motion equation — the method of assuming a power series and determining coefficients so that the differential equation is satisfied. The two independent solutions defined by Eqs. 3.12(8) and 3.12(9) were denoted by $J_0(Tr)$ and $N_0(Tr)$, so that a complete solution to (1) is written

$$R = CJ_0(Tr) + DN_0(Tr) \tag{4}$$

Since the differential equations (1) and (2) are similar, it may be expected that the above solutions are similar to sinusoids, but revised somewhat by the presence of the term $\frac{1}{r}\frac{dR}{dr}$. This is true, for a plot of the two solutions $J_0(v)$ and $N_0(v)$ as functions of v (Fig. 3.15) shows

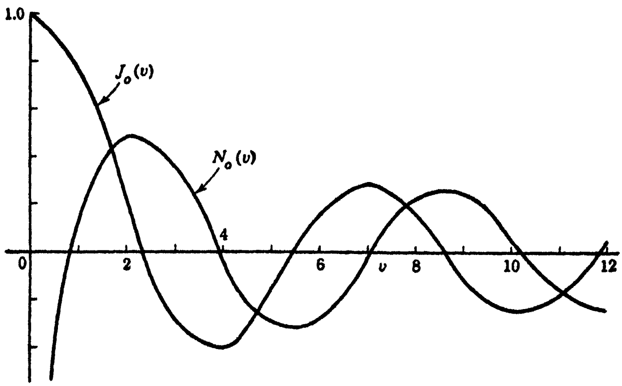


FIG. 3.15. Plot of zero order Bessel functions $J_0(v)$ and $N_0(v)$.

that both are reminiscent of damped sinusoids. $J_0(v)$ is unity at $v = 0$ and then alternates in sign, actually approaching a sinusoidal form as v becomes very large.

$$J_0(v) \underset{v\to\infty}{\longrightarrow} \sqrt{\frac{2}{\pi v}}\cos\left(v - \frac{\pi}{4}\right) \tag{5}$$

$N_0(v)$ is infinite at $v = 0$, but eventually alternates in sign, and for large arguments approaches

$$N_0(v) \underset{v\to\infty}{\longrightarrow} \sqrt{\frac{2}{\pi v}}\sin\left(v - \frac{\pi}{4}\right) \tag{6}$$

Both the functions $J_0(v)$ and $N_0(v)$, called Bessel functions of zero order, first and second kinds respectively, are tabulated extensively in references. Some care should be observed in using these references, for there is a wide variation in notation for the second solution, and not all

the functions used are equivalent since they differ in the values of arbitrary constants selected for the series. The $N_0(v)$ is chosen here to agree with Jahnke and Emde, which now provides the best available tables. It is equivalent to the $Y_0(v)$ used by Watson and by McLachlan, and the $\bar{Y}_0(v)$, but not the $Y_0(v)$, of Gray, Matthews, and MacRobert. Of course, it is quite proper to use any one of the second solutions throughout a given problem, since all the differences will be absorbed in the arbitrary constants of the problem, and the same final numerical result will always be obtained; but it is necessary to be consistent in the use of only one of these throughout any given analysis.

3.16 Linear Combinations of J_0 and N_0: The Hankel Functions

It was found in Chapter 1 that it is sometimes convenient to express the solution to the simple harmonic motion equation in terms of complex exponentials, which are linear combinations of sines and cosines. Thus it is proper to write the solution to Eq. 3.15(2) as

$$Z = A_1 e^{jkz} + B_1 e^{-jkz} \tag{1}$$

since

$$e^{jkz} = \cos kz + j \sin kz$$

and

$$e^{-jkz} = \cos kz - j \sin kz \tag{2}$$

This form is of particular value in the study of traveling waves if $e^{j\omega t}$ is to be used to represent sinusoidal time variations, for then

$$\begin{aligned} Ze^{j\omega t} &= e^{j\omega t}(A_1 e^{jkz} + B_1 e^{-jkz}) \\ &= A_1 e^{j(\omega t + kz)} + B_1 e^{j(\omega t - kz)} \end{aligned}$$

The first of these terms represents a wave traveling in the negative z direction; the second represents a wave traveling in the positive z direction.

Similarly, for the study of wave propagation in cylindrical coordinates, it is convenient to form linear combinations of the Bessel functions $J_0(Tr)$ and $N_0(Tr)$.

$$H_0^{(1)}(Tr) = J_0(Tr) + jN_0(Tr) \tag{3}$$

$$H_0^{(2)}(Tr) = J_0(Tr) - jN_0(Tr) \tag{4}$$

The meaning of these combinations is similar to that of the exponentials. This is shown by substituting the expressions for J_0 and N_0 at large

arguments [Eqs. 3.15(5) and 3.15(6)] in (3) and (4).

$$H_0^{(1)}(v) \underset{v\to\infty}{\longrightarrow} \sqrt{\frac{2}{\pi v}}\left[\cos\left(v - \frac{\pi}{4}\right) + j\sin\left(v - \frac{\pi}{4}\right)\right] = \sqrt{\frac{2}{\pi v}}\,e^{j\left(v-\frac{\pi}{4}\right)} \qquad [5]$$

$$H_0^{(2)}(v) \underset{v\to\infty}{\longrightarrow} \sqrt{\frac{2}{\pi v}}\left[\cos\left(v - \frac{\pi}{4}\right) - j\sin\left(v - \frac{\pi}{4}\right)\right] = \sqrt{\frac{2}{\pi v}}\,e^{-j\left(v-\frac{\pi}{4}\right)} \qquad [6]$$

The solution to Eq. 3.15(1) may be written in terms of the linear combinations defined above.

$$R = C_1H_0^{(1)}(Tr) + D_1H_0^{(2)}(Tr) \qquad [7]$$

If this solution is associated with a time function $e^{j\omega t}$,

$$e^{j\omega t}R = e^{j\omega t}[C_1H_0^{(1)}(Tr) + D_1H_0^{(2)}(Tr)]$$

For large values of Tr,

$$e^{j\omega t}R \longrightarrow \sqrt{\frac{2}{\pi Tr}}\left[C_1e^{-j\frac{\pi}{4}}e^{j(\omega t+Tr)} + D_1e^{j\frac{\pi}{4}}e^{j(\omega t-Tr)}\right]$$

so that the first term represents a wave traveling radially inward and the second term represents a wave traveling radially outward. This interpretation of the functions $H_0^{(1)}$ and $H_0^{(2)}$ will be particularly useful in later chapters concerned with wave solutions.

The functions $H_0^{(1)}(v)$ and $H_0^{(2)}(v)$ are called Hankel functions of the first and second kinds, respectively. However, it should be emphasized that these are not new functions, but merely linear combinations of $J_0(v)$ and $N_0(v)$ and so are also solutions of Bessel's equation. Other complete solutions to the Bessel equation might be written

$$\begin{aligned} R &= C_3J_0(Tr) + D_3H_0^{(1)}(Tr) \\ R &= C_4N_0(Tr) + D_4H_0^{(2)}(Tr), \text{ etc.} \end{aligned} \qquad [8]$$

3.17 Bessel Functions of Zero Order: Imaginary Arguments

If the constant K of the simple harmonic motion equation is imaginary, $K = jk$ or $K^2 = -k^2$ where k is real, the solution, in terms of exponentials, is

$$Z = A_2e^{kz} + B_2e^{-kz}$$

Similarly, the constant T of Bessel's equation is often imaginary: $T = j\tau$ or $T^2 = -\tau^2$ where τ is real. The first solution for this case has been studied in Art. 3.12. The series representing $J_0(j\tau r)$ is real and convergent [see Eq. 3.12(13)] and, as noted in that article, is often

denoted $I_0(\tau r)$, or

$$I_0(v) \equiv J_0(jv)$$

Similarly, the second solution, $N_0(j\tau r)$, could be used if desired, but here it is more convenient to use for the second solution one of the linear combinations of J_0 and N_0 defined in the previous article. This is mainly to facilitate the satisfying of boundary conditions in a region extending to infinity. Consider for instance Eqs. 3.15(5) and 3.15(6) with $v = jv'$. Expansion of these into the hyperbolic function form shows that if v' approaches infinity, both $J_0(jv')$ and $N_0(jv')$ approach infinity.

For the solution in a region extending to infinity the choice of the proper combination of $J_0(jv')$ and $N_0(jv')$ must then be made so that

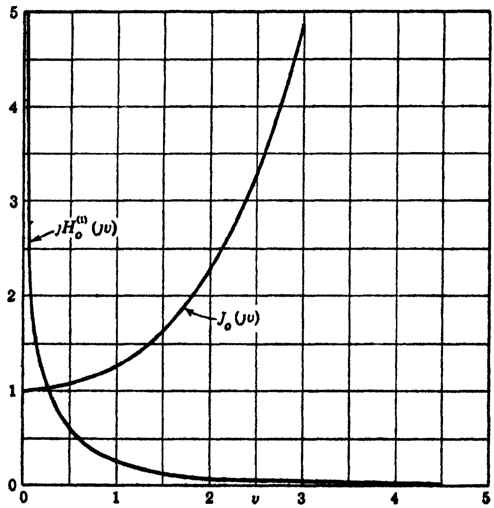


FIG. 3.17. Plot of zero order Bessel functions of imaginary arguments.

the combination of the two does not become infinite for $v' = \infty$; this choice is easily made by noting the linear combination $H_0^{(1)}(v)$ of Eq. 3.16(3) and its value for large arguments, Eq. 3.16(5). For $v = jv'$

$$H_0^{(1)}(jv') \underset{v' \to \infty}{\longrightarrow} \sqrt{\frac{2}{\pi v'}} \frac{1}{\sqrt{j}} e^{-j\frac{\pi}{4}} e^{-v'} = -j\sqrt{\frac{2}{\pi v'}} e^{-v'}$$

This combination of the J and N functions becomes zero properly at infinity, and since it is not linearly dependent upon $J_0(jv)$, it may always be used for the second solution when the argument is imaginary. Thus a

complete solution to the equation

$$\frac{d^2R}{dr^2} + \frac{1}{r}\frac{dR}{dr} - \tau^2 R = 0$$

may always be written

$$R = C_2 J_0(j\tau r) + D_2 H_0^{(1)}(j\tau r)$$

$H_0^{(1)}(jv')$ is always a purely imaginary number (if v' is real), so $jH_0^{(1)}(jv')$ is real. $J_0(jv')$ and $jH_0^{(1)}(jv')$ are tabulated in the references as functions of v', and curves of these functions are given in Fig. 3.17. As may be expected, the former is reminiscent of the hyperbolic cosine, the latter of a negative exponential.

Again there are differences in notation for these solutions in the references. It has been pointed out that $I_0(v')$ is widely used to denote $J_0(jv')$. Similarly, the function

$$K_0(v') = j\frac{\pi}{2}H_0^{(1)}(jv')$$

is used commonly for the second solution, as in Watson, McLachlan, and Gray, Matthews, and MacRobert. I_0 and K_0 are called modified Bessel functions of zero order, first and second kinds respectively.

3.18 Bessel Functions of Higher Order

The simple Bessel equation 3.15(1) was derived by assuming that a product solution would satisfy Laplace's equation, first eliminating any variations with angle ϕ. For certain problems, as, for example, the solution for field between the two halves of a longitudinally split cylinder, it may be necessary to retain the ϕ variations in the equation. The solution may be assumed in product form again, RZF_ϕ, where R is a function of r alone, Z of z alone, and F_ϕ of ϕ alone. Z has solutions in exponentials or sinusoids as before, and F_ϕ may also be satisfied by sinusoids.

$$Z = Ae^{Tz} + Be^{-Tz} \qquad [1]$$

$$F_\phi = E\cos\nu\phi + F\sin\nu\phi \qquad [2]$$

The differential equation for R is then slightly different from the zero order Bessel equation obtained previously.

$$\frac{d^2R}{dr^2} + \frac{1}{r}\frac{dR}{dr} + \left(T^2 - \frac{\nu^2}{r^2}\right)R = 0 \qquad [3]$$

It is apparent at once that Eq. 3.15(1) is a special case of this more general equation, i.e., $\nu = 0$. A series solution to the general equation

carried through as in Art. 3.12 shows that the function defined by the series

$$J_\nu(Tr) = \sum_{m=0}^{\infty} \frac{(-1)^m \left(\frac{Tr}{2}\right)^{\nu+2m}}{m!\,\Gamma(\nu + m + 1)} \qquad [4]$$

is a solution to the equation.

$\Gamma(\nu + m + 1)$ is the gamma function of $(\nu + m + 1)$ and for ν integral is equivalent to the factorial of $(\nu + m)$; for ν non-integral, values of this gamma function are tabulated. If ν is an integer n,

$$J_n(Tr) = \sum_{m=0}^{\infty} \frac{(-1)^m \left(\frac{Tr}{2}\right)^{n+2m}}{m!\,(n + m)!} \qquad [5]$$

Similarly, a second independent solution[1] to the equation is

$$N_\nu(Tr) = \frac{\cos \nu\pi J_\nu(Tr) - J_{-\nu}(Tr)}{\sin \nu\pi} \qquad [6]$$

So that a complete solution to (3) may be written

$$R = AJ_\nu(Tr) + BN_\nu(Tr) \qquad [7]$$

The constant ν is known as the order of the equation. J_ν is then called a Bessel function of first kind, order ν; N_ν is a Bessel function of second kind, order ν. Of most interest for this chapter are cases in which $\nu = n$, an integer.

The solution to (3) may also be written in terms of linear combinations of J_ν and N_ν

$$R = A_1H_\nu^{(1)}(Tr) + B_1H_\nu^{(2)}(Tr) \qquad [8]$$

where

$$H_\nu^{(1)}(Tr) = J_\nu(Tr) + jN_\nu(Tr) \qquad [9]$$

and

$$H_\nu^{(2)}(Tr) = J_\nu(Tr) - jN_\nu(Tr) \qquad [10]$$

$H_\nu^{(1)}$ and $H_\nu^{(2)}$ are called Hankel functions of order ν, first and second kinds respectively.

If T is imaginary, $T = jr$, (3) becomes

$$\frac{d^2R}{dr^2} + \frac{1}{r}\frac{dR}{dr} - \left(r^2 + \frac{\nu^2}{r^2}\right)R = 0 \qquad [11]$$

[1] If ν is non-integral, $J_{-\nu}$ is not linearly related to J_ν, and it is then proper to use either $J_{-\nu}$ or N_ν as the second solution; for ν integral, N_ν must be used. Equation (6) is indeterminate for ν integral but is subject to evaluation by usual methods.

If $T = j\tau$ is substituted in the series definitions for $J_\nu(Tr)$ and $H_\nu^{(1)}(Tr)$, the resulting quantities are found to be either pure reals or pure imaginaries if ν is an integer, n. Specifically, the quantities $[j^{-n}J_n(jv)]$ and $[j^{n+1}H_n^{(1)}(jv)]$ are always pure real numbers and are tabulated as functions of v in the references. A complete solution to (11) may be written

$$R = A_2 J_\nu(j\tau r) + B_2 H_\nu^{(1)}(j\tau r) \tag{12}$$

Again it is quite common practice to denote the above solutions as follows.

$$I_n(v) = j^{-n} J_n(jv) \tag{13}$$

$$K_n(v) = \frac{\pi}{2} j^{n+1} H_n^{(1)}(jv) \tag{14}$$

The solution to (11) may then also be written

$$R = A_3 I_\nu(\tau r) + B_3 K_\nu(\tau r) \tag{15}$$

However, note that I_ν and K_ν as defined by (13) and (14) will not always satisfy the recurrence formulas given for general Bessel functions in following articles.

3.19 Values for Bessel Functions of Large Arguments

As the arguments of the Bessel functions become very large, all these functions approach more and more closely sinusoidal or exponential forms, as in the zero order functions of Arts. 3.15–3.17. These forms are called the asymptotic expressions for the Bessel functions. They are as follows:

$$\underset{v \to \infty}{J_\nu(v)} \to \sqrt{\frac{2}{\pi v}} \cos\left(v - \frac{\pi}{4} - \frac{\nu\pi}{2}\right) \tag{1}$$

$$\underset{v \to \infty}{N_\nu(v)} \to \sqrt{\frac{2}{\pi v}} \sin\left(v - \frac{\pi}{4} - \frac{\nu\pi}{2}\right) \tag{2}$$

$$\underset{v \to \infty}{H_\nu^{(1)}(v)} \to \sqrt{\frac{2}{\pi v}}\, e^{j\left(v - \frac{\pi}{4} - \frac{\nu\pi}{2}\right)} \tag{3}$$

$$\underset{v \to \infty}{H_\nu^{(2)}(v)} \to \sqrt{\frac{2}{\pi v}}\, e^{-j\left(v - \frac{\pi}{4} - \frac{\nu\pi}{2}\right)} \tag{4}$$

$$\underset{v \to \infty}{j^{-\nu} J_\nu(jv)} = \underset{v \to \infty}{I_\nu(v)} \to \sqrt{\frac{1}{2\pi v}}\, e^{v} \tag{5}$$

$$\underset{v \to \infty}{j^{\nu+1} H_\nu^{(1)}(jv)} = \frac{2}{\pi} \underset{v \to \infty}{K_\nu(v)} \to \sqrt{\frac{2}{\pi v}}\, e^{-v} \tag{6}$$

3.20 Differentiation of Bessel Functions

If it is desired to obtain the differential of the Bessel function $J_0(v)$, the series definition 3.12(8) may be differentiated term by term.

$$\frac{d}{dv}[J_0(v)] = \frac{d}{dv}\left[1 - \left(\frac{v}{2}\right)^2 + \frac{\left(\frac{v}{2}\right)^4}{(2!)^2} - \frac{\left(\frac{v}{2}\right)^6}{(3!)^2} + \frac{\left(\frac{v}{2}\right)^8}{(4!)^2} - \cdots\right]$$

$$= -\left[\frac{v}{2} - \frac{\left(\frac{v}{2}\right)^3}{1 \cdot 2!} + \frac{\left(\frac{v}{2}\right)^5}{2!\,3!} - \frac{\left(\frac{v}{2}\right)^7}{3!\,4!} + \cdots\right]$$

Comparison with the definition for $J_n(v)$, Eq. 3.18(5), shows that the result is exactly $-J_1(v)$. That is,

$$\frac{d}{dv}[J_0(v)] = -J_1(v)$$

Similarly, it may be shown from the series definitions that the following derivative expressions are true for any of the Bessel functions $J_\nu(v)$, $N_\nu(v)$, $H_\nu^{(1)}(v)$, or $H_\nu^{(2)}(v)$. Let $R_\nu(v)$ denote any one of these, and R_ν' denote $(d/dv)[R_\nu(v)]$.

$$R_0'(v) = -R_1(v) \qquad [1]$$

$$R_1'(v) = R_0(v) - \frac{1}{v}R_1(v) \qquad [2]$$

$$vR_\nu'(v) = \nu R_\nu(v) - vR_{\nu+1}(v) \qquad [3]$$

$$vR_\nu'(v) = -\nu R_\nu(v) + vR_{\nu-1}(v) \qquad [4]$$

$$\frac{d}{dv}[v^{-\nu}R_\nu(v)] = -v^{-\nu}R_{\nu+1}(v) \qquad [5]$$

$$\frac{d}{dv}[v^{\nu}R_\nu(v)] = v^{\nu}R_{\nu-1}(v) \qquad [6]$$

Note that

$$R_\nu'(Tr) = \frac{d}{d(Tr)}[R_\nu(Tr)] = \frac{1}{T}\frac{d}{dr}[R_\nu(Tr)] \qquad [7]$$

For the I and K functions different forms for the above differentials must be used. They may be obtained from the above by substituting

Eqs. 3.18(13) and 3.18(14) in the preceding expressions. Some of these are

$$\begin{aligned} vI'_\nu(v) &= \nu I_\nu(v) + vI_{\nu+1}(v) \\ vI'_\nu(v) &= -\nu I_\nu(v) + vI_{\nu-1}(v) \end{aligned} \tag{8}$$

$$\begin{aligned} vK'_\nu(v) &= \nu K_\nu(v) - vK_{\nu+1}(v) \\ vK'_\nu(v) &= -\nu K_\nu(v) - vK_{\nu-1}(v) \end{aligned} \tag{9}$$

3.21 Recurrence Formulas for Bessel Functions

By recurrence formulas, it is possible to obtain the value for Bessel functions of any order, when the values of functions for any two other orders, differing from the first by integers, are known. For example, subtract Eq. 3.20(4) from Eq. 3.20(3). The result may be written

$$\frac{2\nu}{v} R_\nu(v) = R_{\nu+1}(v) + R_{\nu-1}(v) \tag{1}$$

By this equation, if any two of $R_{\nu-1}$, R_ν, and $R_{\nu+1}$ are known, the third may be found. For instance, if $J_0(v)$ and $J_1(v)$ are known, the equation may be used to find $J_2(v)$; repeating the process with $J_1(v)$ and $J_2(v)$, $J_3(v)$ may be determined, and so on to any order desired. As before, R_ν may denote J_ν, N_ν, $H_\nu^{(1)}$, $H_\nu^{(2)}$ but not I_ν or K_ν. For these, the recurrence formulas are

$$\frac{2\nu}{v} I_\nu(v) = I_{\nu-1}(v) - I_{\nu+1}(v) \tag{2}$$

$$\frac{2\nu}{v} K_\nu(v) = K_{\nu+1}(v) - K_{\nu-1}(v) \tag{3}$$

3.22 Integrals of Bessel Functions

Equation 3.20(1) may be integrated directly

$$\int R_1(v)\, dv = -R_0(v) \tag{1}$$

Others of the integrals that will be useful in solving later problems are given below. R_ν denotes J_ν, N_ν, $H_\nu^{(1)}$, or $H_\nu^{(2)}$.

$$\int v^{-\nu} R_{\nu+1}(v)\, dv = -v^{-\nu} R_\nu(v) \tag{2}$$

$$\int v^{\nu} R_{\nu-1}(v)\, dv = v^{\nu} R_\nu(v) \tag{3}$$

$$\int vR_\nu(\alpha v)R_\nu(\beta v)\,dv$$

$$= \frac{v}{\alpha^2 - \beta^2}[\beta R_\nu(\alpha v)R_{\nu-1}(\beta v) - \alpha R_{\nu-1}(\alpha v)R_\nu(\beta v)] \qquad \alpha \neq \beta \qquad [4]$$

$$\int vR_\nu^2(\alpha v)\,dv = \frac{v^2}{2}[R_\nu^2(\alpha v) - R_{\nu-1}(\alpha v)R_{\nu+1}(\alpha v)]$$

$$= \frac{v^2}{2}\left[R_\nu'^2(\alpha v) + \left(1 - \frac{\nu^2}{\alpha^2 v^2}\right)R_\nu^2(\alpha v)\right] \qquad [5]$$

3.23 Expansion of a Function as a Series of Bessel Functions

In Chapter 1 a study was made of the familiar method of Fourier series by which a function may be expressed over a given region as a series of sines or cosines. It is possible to evaluate the coefficients in such a case because of the orthogonality property of sinusoids, expressed in Art. 1.13. A study of the integrals Eqs. 3.22(4) and 3.22(5) shows that there are similar orthogonality expressions for Bessel functions. For example, these integrals may be written for zero order Bessel functions, and if α and β are taken as p_m/a and p_q/a, where p_m and p_q are the mth and qth roots of $J_0(v) = 0$, that is, $J_0(p_m) = 0$ and $J_0(p_q) = 0$, $p_m \neq p_q$, then Eq. 3.22(4) gives

$$\int_0^a rJ_0\left(\frac{p_m r}{a}\right)J_0\left(\frac{p_q r}{a}\right)dr = 0, \qquad p_m \neq p_q \qquad [1]$$

So a function $f(r)$ may be expressed as an infinite sum of zero order Bessel functions.

$$f(r) = b_1J_0\left(p_1\frac{r}{a}\right) + b_2J_0\left(p_2\frac{r}{a}\right) + b_3J_0\left(p_3\frac{r}{a}\right) + \cdots$$

or

$$f(r) = \sum_{m=1}^{\infty} b_mJ_0\left(\frac{p_m r}{a}\right) \qquad [2]$$

The coefficients b_m may be evaluated in a manner similar to that used for Fourier coefficients by multiplying each term of (2) by $rJ_0(p_m r/a)$ and integrating from zero to a. Then by (1) all terms on the right disappear except the mth term.

$$\int_0^a rf(r)J_0\left(\frac{p_m r}{a}\right)dr = \int_0^a b_m r\left[J_0\left(\frac{p_m r}{a}\right)\right]^2 dr$$

From Eq. 3.22(5)

$$\int_0^a rJ_0^2\left(\frac{p_m r}{a}\right) dr = \frac{a^2}{2} J_1^2(p_m) \qquad [3]$$

So

$$\int_0^a rf(r)J_0\left(\frac{p_m r}{a}\right) dr = \frac{b_m a^2}{2} J_1^2(p_m)$$

or

$$b_m = \frac{2}{a^2 J_1^2(p_m)} \int_0^a rf(r)J_0\left(\frac{p_m r}{a}\right) dr \qquad [4]$$

Thus the coefficients for the series (2) are determined. Mathematical study shows that the series is convergent and properly expresses any decently behaved function $f(r)$ over the region $r = 0$ to $r = a$.

Problem 3.23(*a*). Write a function $f(r)$ in terms of nth order Bessel functions over the range 0 to a and determine the coefficients.

Problem 3.23(*b*). Determine coefficients for a function $f(r)$ expressed over the range 0 to a as a series of zero order Bessel functions as follows.

$$f(r) = \sum_{m=1}^{\infty} c_m J_0\left(\frac{p'_m r}{a}\right)$$

where p'_m denotes the mth root of $J'_0(v) = 0$ [i.e., $J_1(v) = 0$].

3.24 Cylindrical Harmonic Series for Radial Matching

If it is desired to find potential distribution in the interior of a region bounded by a circular cylinder and its base at potential zero, and a plane perpendicular to the axis at potential V_0 (Fig. 3.24*a*), the use of series is similar to that of the example in Art. 3.14, although now a series of Bessel functions is required. The gaps are again assumed negligibly small compared with all other dimensions. In selecting the proper form for the solution from Art. 3.12, the boundary condition that $\Phi = 0$ at $r = a$ for all values of z indicates that the R function must become zero at $r = a$. Thus we select the J_0 functions since the I_0's do not ever become zero. (The corresponding second solution, N_0, does not appear since potential must remain finite on the axis.) The value of T in Eq. 3.12(10) is determined from the condition that $\Phi = 0$ at $r = a$ for all values of z. Thus if p_m is the mth root of $J_0(v) = 0$, T must be p_m/a. The corresponding solution for Z is in hyperbolic functions, Eq. 3.12(11), but the coefficient of the hyperbolic cosine term must be zero since Φ is zero at $z = 0$ for all values of r. Thus a sum of all cylindrical harmonics with arbitrary amplitudes which satisfy the sym-

metry of the problem and the boundary conditions so far imposed may be written

$$\Phi = \sum_{m=1}^{\infty} B_m J_0\left(\frac{p_m r}{a}\right) \sinh\left(\frac{p_m z}{a}\right) \quad [1]$$

The remaining condition is that at $z = l$, $\Phi = 0$ at $r = a$ and $\Phi = V_0$ for all other r's. To use this condition it seems advisable to expand such

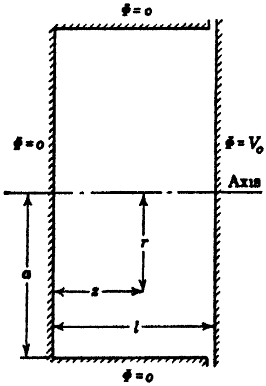


FIG. 3.24a. Cylindrical region ($l/a = 1$).

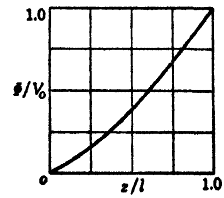


FIG. 3.24b. Plot of potential vs. z at $r = 0$ ($l/a = 1$).

a function over this plane in terms of Bessel functions as in Eq. 3.23(2). For the coefficients, Eq. 3.23(4) is used with $f(r) = 0$ at $r = a$ and $f(r) = V_0$ for $0 < r < a$. Then

$$b_m = \frac{2}{a^2 J_1^2(p_m)} \int_0^a r V_0 J_0\left(\frac{p_m r}{a}\right) dr = \frac{2V_0}{p_m J_1(p_m)}$$

The above integral was evaluated by Eq. 3.22(3). So

$$f(r) = \Phi_{z=l} = \sum_{m=1}^{\infty} \frac{2V_0}{p_m J_1(p_m)} J_0\left(\frac{p_m r}{a}\right) \quad [2]$$

But (1) at $z = l$ is

$$\Phi_{z=l} = \sum_{m=1}^{\infty} B_m \sinh\left(\frac{p_m l}{a}\right) J_0\left(\frac{p_m r}{a}\right) \quad [3]$$

Equations (2) and (3) must be equivalent for all values of r. Consequently, coefficients of corresponding terms of $J_0(p_m r/a)$ must be equal. The constant B_m is now completely determined, and the potential at any

point inside the region is

$$\Phi = \sum_{m=1}^{\infty} \frac{2V_0}{p_m J_1(p_m) \sinh\left(\dfrac{p_m l}{a}\right)} \sinh\left(\frac{p_m z}{a}\right) J_0\left(\frac{p_m r}{a}\right) \qquad [4]$$

Potential distribution along the axis is plotted in Fig. 3.24*b* for a case with $a/l = 1$.

3.25 Determination of Amplitudes of Cylindrical Harmonics in General

The specific examples of the preceding articles are special cases of more general problems. That is, Art. 3.14 is a special case of the problem: desired the potential distribution inside a region bounded by a cylindrical surface over which potential is given as $\Phi = f(z)$ at $r = r_0$, and two zero-potential planes perpendicular to the axis and distance l apart.

Then

$$\Phi = \sum_{m=1}^{\infty} b_m \frac{I_0\left(\dfrac{m\pi r}{l}\right)}{I_0\left(\dfrac{m\pi r_0}{l}\right)} \sin\left(\frac{m\pi z}{l}\right) \qquad [1]$$

where

$$b_m = \frac{2}{l}\int_0^l f(z) \sin\frac{m\pi z}{l}\, dz \qquad [2]$$

Article 3.24 was a special case of the problem: desired the potential distribution inside a region bounded by a cylinder of radius r_0 and its plane base at zero potential, and a second plane surface perpendicular to the axis. Potential is given over the latter plane surface as $\Phi = f(r)$ at $z = l$; then

$$\Phi = \sum_{m=1}^{\infty} d_m \frac{\sinh\left(\dfrac{p_m z}{r_0}\right)}{\sinh\left(\dfrac{p_m l}{r_0}\right)} J_0\left(\frac{p_m r}{r_0}\right) \qquad [3]$$

where

$$d_m = \frac{2}{r_0^2 J_1^2(p_m)} \int_0^{r_0} r f(r) J_0\left(\frac{p_m r}{r_0}\right) dr \qquad [4]$$

p_m is the mth root of $J_0(v) = 0$.

The series solutions obtained for three-dimensional problems actually have an infinite number of terms in general, and the method would be of little use if it were necessary to calculate a great many of these terms for practical problems. The actual number that must be included will depend upon the desired accuracy and the irregularity of the boundaries. Often it is only necessary to include three or four terms to give a very satisfactory representation of a physical problem. The curves plotted in Figs. 3.14c and 3.24b were calculated, using from three to six terms of the series.

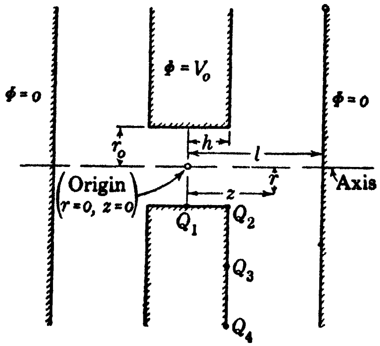


FIG. 3.25a. Region of cylindrical symmetry $\left(\frac{r_0}{l} = \frac{h}{l} = 0.3\right)$.

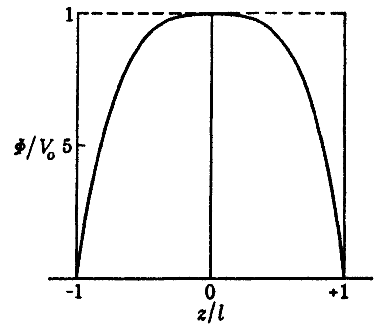


FIG. 3.25b. Plot of potential on axis.

In many three-dimensional regions over which field distribution is required, potentials are not known over any single simple surface, so then the integrals (2) and (4) cannot be evaluated to give the coefficients of the series of product solutions. Still it is often possible to describe field distribution accurately enough in such a region by retaining only three or four terms of the series, determining coefficients by selecting as many points at which potential is known, and solving simultaneously for the coefficients. An example will clarify the procedure.

Consider an electrostatic electron lens made up of a thick metal electrode with a round hole, midway between two other plane electrodes. The inner electrode is at potential V_0 and the outer two are at zero potential. The lens has cylindrical symmetry about an axis, as shown by Fig. 3.25a, so that a series of cylindrical harmonics from Art. 3.12 will describe the field. Potential must become zero at $z = \pm l$ if the origin is taken at the center; the distribution should be an even, not an odd function, so only cosine terms can be present. Retaining four terms of the series, potential inside the region may then be described approxi-

mately by (Art. 3.12)

$$\frac{\Phi}{V_0} = a_1 I_0\left(\frac{\pi r}{2l}\right)\cos\left(\frac{\pi z}{2l}\right) + a_2 I_0\left(\frac{3\pi r}{2l}\right)\cos\left(\frac{3\pi z}{2l}\right) + a_3 I_0\left(\frac{5\pi r}{2l}\right)\cos\left(\frac{5\pi z}{2l}\right) + a_4 I_0\left(\frac{7\pi r}{2l}\right)\cos\left(\frac{7\pi z}{2l}\right) \quad [5]$$

The four constants, a_1, a_2, a_3, a_4, may be determined by selecting any four points, as Q_1, Q_2, Q_3, Q_4, on the inner electrode where $\Phi = V_0$, and solving the four resulting equations simultaneously for four a's. In the example shown, $r_0/l = h/l = 0.3$. The four points selected to determine the four constants were

1. $\dfrac{\Phi}{V_0} = 1$ at $z = 0, \quad r = r_0$
2. $\dfrac{\Phi}{V_0} = 1$ at $z = h, \quad r = r_0$
3. $\dfrac{\Phi}{V_0} = 1$ at $z = h, \quad r = 1.5r_0$
4. $\dfrac{\Phi}{V_0} = 1$ at $z = h, \quad r = 3r_0$

These values substituted in (5) give four separate equations which may be solved simultaneously to give:

$$a_1 = 1.15 \quad a_2 = -0.172 \quad a_3 = 0.0211 \quad a_4 = 0.00098$$

From (5), with these values substituted, potential may be found at any point r, z. In particular, a plot of distribution along the axis is shown in Fig. 3.25*b*.

3.26 Spherical Harmonics

Consider next a problem whose configuration suggests spherical coordinates, as for example the potential distribution due to two thin hemispherical shells at different potentials separated by a gap negligibly small compared with the radius (Fig. 3.26*a*). Laplace's equation in spherical coordinates is (Art. 2.38)

$$\frac{1}{r}\frac{\partial^2(r\Phi)}{\partial r^2} + \frac{1}{r^2\sin\theta}\frac{\partial\left(\sin\theta\dfrac{\partial\Phi}{\partial\theta}\right)}{\partial\theta} + \frac{1}{r^2\sin^2\theta}\frac{\partial^2\Phi}{\partial\phi^2} = 0$$

Axial symmetry eliminates variations with ϕ, so

$$\frac{\partial^2(r\Phi)}{\partial r^2} + \frac{1}{r \sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\Phi}{\partial\theta}\right) = 0 \qquad [1]$$

or

$$r\frac{\partial^2\Phi}{\partial r^2} + 2\frac{\partial\Phi}{\partial r} + \frac{1}{r}\frac{\partial^2\Phi}{\partial\theta^2} + \frac{1}{r\tan\theta}\frac{\partial\Phi}{\partial\theta} = 0 \qquad [2]$$

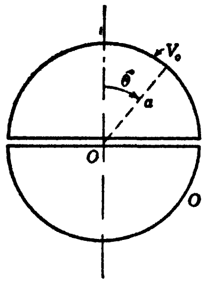


FIG. 3.26*a*. Two hemispheres separated by a gap.

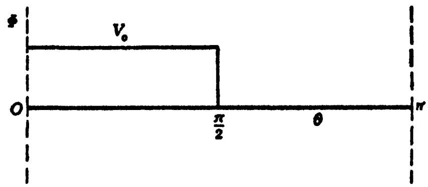


FIG. 3.26*b*. Plot of potential vs. θ at $r = a$.

Assume a product solution,

$$\Phi = R\Theta$$

where R is a function of r alone, Θ of θ alone,

$$rR''\Theta + 2R'\Theta + \frac{1}{r}R\Theta'' + \frac{1}{r\tan\theta}R\Theta' = 0$$

and

$$\frac{r^2R''}{R} + \frac{2rR'}{R} = -\frac{\Theta''}{\Theta} - \frac{\Theta'}{\Theta\tan\theta} \qquad [3]$$

Following previous logic, if the two sides of the equations are to be equal to each other for all values of r and θ, both sides can be equal only to a constant. Since the constant may be expressed in any non-restrictive way, let it be $m(m + 1)$. The two resulting ordinary differential equations are then

$$r^2\frac{d^2R}{dr^2} + 2r\frac{dR}{dr} - m(m+1)R = 0 \qquad [4]$$

$$\frac{d^2\Theta}{d\theta^2} + \frac{1}{\tan\theta}\frac{d\Theta}{d\theta} + m(m+1)\Theta = 0 \qquad [5]$$

The first of these has a solution which is easily verified to be

$$R = C_1r^m + C_2r^{-(m+1)} \qquad [6]$$

A solution to the second equation in terms of simple functions is not obvious, so as with the Bessel equation, a series solution may be assumed. The coefficients of this series must be determined so that the differential equation (5) is satisfied and the resulting series made to define a new function. There is one departure here from an exact analogue with the Bessel functions, for it turns out that a proper selection of the arbitrary constants will make the series for the new function terminate in a finite number of terms if m is an integer. Thus for any integer m, the polynomial defined by

$$P_m(\cos\theta) \equiv \frac{1}{2^m m!}\left[\frac{d}{d(\cos\theta)}\right]^m (\cos^2\theta - 1)^m \tag{7}$$

is a solution to the differential equation (5). The equation is known as Legendre's equation; the solutions are called Legendre polynomials of order m. Their forms for the first few values of m are tabulated below. It is evident that since they are polynomials and not infinite series, their values can be calculated exactly if desired, but values of the polynomials are also tabulated in many references.

$$\begin{aligned}
P_0(\cos\theta) &= 1\\
P_1(\cos\theta) &= \cos\theta\\
P_2(\cos\theta) &= \tfrac{1}{2}(3\cos^2\theta - 1)\\
P_3(\cos\theta) &= \tfrac{1}{2}(5\cos^3\theta - 3\cos\theta)\\
P_4(\cos\theta) &= \tfrac{1}{8}(35\cos^4\theta - 30\cos^2\theta + 3)\\
P_5(\cos\theta) &= \tfrac{1}{8}(63\cos^5\theta - 70\cos^3\theta + 15\cos\theta)
\end{aligned} \tag{8}$$

It is recognized that $\Theta = C_1 P_m(\cos\theta)$ is only one solution to the second-order differential equation (5). There must be a second independent solution, which may be obtained in a similar manner, but it turns out that this solution becomes infinite for $\theta = 0$. Consequently it will never be present for any case in which the axis of spherical coordinates is included in the region over which the solution applies. However, several important situations require their use; when this occurs certain of the references should be consulted.

Returning to the solution for the equation in r, given by (6), we see that if field inside the shell is desired, the constant C_2 must be zero, since potential cannot become infinite at $r = 0$. If field outside the shell is desired, the constant C_1 must be zero, since potential cannot become infinite at $r = \infty$. Thus all spherical harmonics (as these particular product solutions to Laplace's equation are called) which satisfy the symmetry of the problem of Fig. 3.26a and the boundary conditions so

far imposed may be written

$$\Phi_{\text{inside}} = \sum_{m=0}^{\infty} A_m r^m P_m(\cos\theta) \tag{9}$$

$$\Phi_{\text{outside}} = \sum_{m=0}^{\infty} B_m r^{-(m+1)} P_m(\cos\theta) \tag{10}$$

The additional boundary condition remains that at $r = a$, $\Phi = V_0$ for $0 < \theta < \pi/2$; $\Phi = 0$ for $\pi/2 < \theta < \pi$ for all values of ϕ. If this $f(\theta)$ is plotted against θ, it has the form of Fig. 3.26*b* and could be expanded in a Fourier series over the region $0 < \theta < \pi$, although in order to compare term for term between such a Fourier series and the series (9) or (10), it is desirable rather to express the square wave of Fig. 3.26*b* in terms of Legendre polynomials directly. An orthogonality relation for these polynomials is quite similar to those for sinusoids and Bessel functions which led to the Fourier series and expansion in Bessel functions respectively.

$$\int_0^\pi P_m(\cos\theta)P_n(\cos\theta)\sin\theta\, d\theta = 0, \quad m \neq n \tag{11}$$

It follows that if $f(\theta)$ is defined between the limits of 0 to π, it may be written

$$\begin{aligned} f(\theta) &= \alpha_0 P_0(\cos\theta) + \alpha_1 P_1(\cos\theta) + \alpha_2 P_2(\cos\theta) + \cdots \\ &= \sum_{m=0}^{\infty} \alpha_m P_m(\cos\theta) \end{aligned} \tag{12}$$

where

$$\alpha_m = \frac{2m+1}{2}\int_0^\pi f(\theta)P_m(\cos\theta)\sin\theta\, d\theta \tag{13}$$

For the present problem,

$$\begin{aligned} f(\theta) &= V_0 \quad \text{for} \quad 0 < \theta < \frac{\pi}{2} \\ &= 0 \quad \text{for} \quad \frac{\pi}{2} < \theta < \pi \end{aligned} \tag{14}$$

An integration of (13) for such a $f(\theta)$ would yield

$$f(\theta) = \Phi\bigg|_{r=a} = V_0\left[\frac{1}{2} + \frac{3}{4}P_1(\cos\theta) - \frac{7}{8}\cdot\frac{1}{2}P_3(\cos\theta) + \frac{11}{12}\cdot\frac{1\cdot 3}{2\cdot 4}P_5(\cos\theta)\cdots\right] \tag{15}$$

But (9) gives

$$\Phi_{\text{inside}}\Big|_{r=a} = A_0 + A_1 a P_1(\cos\theta) + A_2 a^2 P_2(\cos\theta) + \cdots \qquad [16]$$

These two expressions must be identical for all values of θ; consequently, they may be equated term by term and all A_m's evaluated. The potential at any point inside the shell is then given by the series

$$\Phi_{\text{inside}} = V_0\left[\frac{1}{2} + \frac{3}{4}\frac{r}{a}P_1(\cos\theta) - \frac{7}{8}\cdot\frac{1}{2}\frac{r^3}{a^3}P_3(\cos\theta) + \cdots\right] \qquad [17]$$

Similarly, the series giving potential at any point outside the shell is found to be

$$\Phi_{\text{outside}} = V_0\left[\frac{a}{2r} + \frac{3}{4}\frac{a^2}{r^2}P_1(\cos\theta) - \frac{7}{8}\cdot\frac{1}{2}\frac{a^4}{r^4}P_3(\cos\theta) + \cdots\right] \qquad [18]$$

3.27 Description of a Field in Spherical Harmonics When the Potential Is Specified on a Spherical Boundary

The example of the previous section was a special case of the more general problem: desired the potential distribution when the axially symmetric potential is defined everywhere on a spherical surface of radius a, by $\Phi\big|_{r=a} = f(\theta)$. Then

$$\Phi_{\text{inside}} = \sum_{m=0}^{\infty} \frac{\alpha_m r^m}{a^m} P_m(\cos\theta) \qquad [1]$$

$$\Phi_{\text{outside}} = \sum_{m=0}^{\infty} \frac{\alpha_m a^{m+1}}{r^{m+1}} P_m(\cos\theta) \qquad [2]$$

where

$$\alpha_m = \frac{2m+1}{2}\int_0^{\pi} f(\theta)P_m(\cos\theta)\sin\theta\, d\theta \qquad [3]$$

3.28 Expansion in Spherical Harmonics When Field Is Given Along an Axis

It is often relatively simple to obtain the field or potential along an axis of symmetry by direct application of fundamental laws, yet difficult to obtain it at any point off this axis by the same technique. Once field is found along an axis of symmetry, expansions in spherical harmonics give its value at any other point. Thus if potential, or any component of field which satisfies Laplace's equation, is given for every point along an axis in such a form that it may be expanded in a power series in z,

the distance along this axis,

$$\Phi\Big|_{\text{axis}} = \sum_{m=0}^{\infty} b_m z^m \tag{1}$$

(power series good for $0 < z < a$)

this axis may be taken as the axis of spherical coordinates, $\theta = 0$, so that z then coincides with r for $\theta = 0$. Equation 3.26(9) gives the potential distribution for a region containing $r = 0$. Along the axis $\theta = 0$, all $P_m(\cos\theta)$ become unity so that

$$\Phi\Big|_{\theta=0} = \sum_{m=0}^{\infty} A_m r^m$$

A comparison with (1) shows that A_m must be equal to the known b_m, so that potential is given at any point inside a sphere of radius a by the series

$$\Phi = \sum_{m=0}^{\infty} b_m r^m P_m(\cos\theta) \tag{2}$$

If potential is desired off the axis outside of this region, the potential along the axis must be expanded in a power series good for $a < z < \infty$.

$$\Phi\Big|_{\theta=0} = \sum_{m=1}^{\infty} c_m z^{-(m+1)}, \quad z > a \tag{3}$$

Then Φ at any point outside is given by comparison with Eq. 3.26(10).

$$\Phi = \sum_{m=0}^{\infty} c_m P_m(\cos\theta) r^{-(m+1)}, \quad r > a \tag{4}$$

3.29 Magnetic Field of Helmholtz Coils at a Point Off the Axis

As an example of the use of the method of Art. 3.26, let us calculate the uniformity of magnetic field over a region around the axis of two large coils, placed as shown in Fig. 3.29. Coil cross sections are negligibly small compared with coil diameters. On the axis, magnetic field has an axial component only, and is given simply from Ampère's law in a manner similar to that of Art. 2.35.

$$\begin{aligned} H_z &= 2\pi a^2 I\left[\frac{1}{[a^2 + (d+z)^2]^{3/2}} + \frac{1}{[a^2 + (d-z)^2]^{3/2}}\right] \\ &= \frac{2\pi a^2 I}{(a^2 + d^2)^{3/2}}\left[\left(1 + \frac{z(z+2d)}{a^2+d^2}\right)^{-3/2} + \left(1 + \frac{z(z-2d)}{a^2+d^2}\right)^{-3/2}\right] \end{aligned} \tag{1}$$

The binomial expansion

$$(1+u)^{-3/2} = 1 - \tfrac{3}{2}u + \tfrac{15}{8}u^2 - \tfrac{105}{48}u^3 + \cdots$$

is good for $0 < u < 1$. Applied to (1)

$$H_z = \frac{2\pi a^2 I}{(a^2 + d^2)^{3/2}} [2 + A_2 z^2 + A_4 z^4 + \cdots]$$

where

$$A_2 = \frac{15d^2 - 3(a^2 + d^2)}{(a^2 + d^2)^2}$$

$$A_4 = \frac{15(a^2 + d^2)^2 - 210d^2(a^2 + d^2) + 315d^4}{4(a^2 + d^2)^4}$$

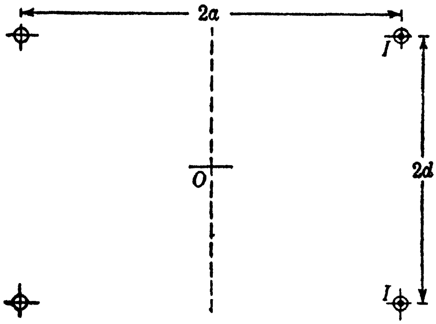


FIG. 3.29. Cross section of Helmholtz coils of large radius.

Since H_z, axial component of magnetic field, satisfies Laplace's equation (Art. 3.02), H_z at any point off the axis is given by

$$H_z = \frac{2\pi a^2 I}{(a^2 + d^2)^{3/2}} [2 + A_2 P_2(\cos\theta) r^2 + A_4 P_4(\cos\theta) r^4 + \cdots] \quad [2]$$

For a region very near the origin (r small) the first terms will be the most important, so field will be most uniform over this region for the condition making $A_2 = 0$. This condition is $a = 2d$ as seen for the expression of A_2 above.

An attempt to obtain H_z at points not on the axis from Ampère's law directly would lead to expressions very difficult to integrate.

3.30 Theory of the Image Method

Point Image in a Plane. The method of images is useful when it is desired to find the field arising from point charges or line charges in the vicinity of conductors. The method is suggested by the simplest case, that of a point charge near a grounded conducting plane (Fig. 3.30a).

Boundary conditions require that the potential along the plane be zero. The requirement is met if in place of the conducting plane, an

equal and opposite image charge is placed at $x = -d$. Potential at any point P is then given by

$$\Phi_P = q\left(\frac{1}{r} - \frac{1}{r'}\right) \qquad [1]$$

This reduces to the required zero potential along the plane since there $r = r'$. Equation (1) then gives the actual potential at any point P to the right of the plane when the charge q is at $x = d$. If the plane is at

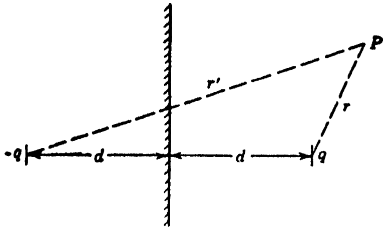


FIG. 3.30*a*. Image of a charge in a conducting plane.

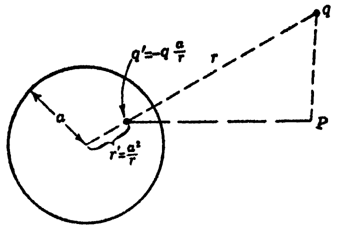


FIG. 3.30*b*. Image of a point charge in a conducting sphere.

any other potential than zero it is necessary only to superpose an additional charge on the plane to give the additional potential. The above equation is, of course, not correct for potential to the left of the plane, since potential is zero there.

Point Image in a Sphere. Potential distribution due to a point charge in front of a conducting sphere can also be solved by an image method, but here the imaging is not so obvious. It can be checked to show that if a charge q is located at a distance r from the center of a conducting sphere of radius a (Fig. 3.30*b*), the required constant potential condition along the boundary of the conducting sphere is satisfied if an image charge of $(-qa/r)$ is placed on the radial line to q at a distance (a^2/r) from the center. The potential at any point P outside the sphere can then be calculated from the charge and its image in the absence of the conducting sphere.

Line Images. Field distribution due to a line charge placed near a conducting plane is found from the line charge and its image in a manner exactly similar to that for the point charge near the plane. For a line charge near a conducting cylinder and parallel to its axis, there is also an image method. It can be checked to show that for a line charge λ per unit length parallel to the axis of a conducting circular cylinder, and a distance r from the axis, the required condition of constant potential over the surface of the cylinder is satisfied by the line charge and an image $-\lambda$ placed at a distance (a^2/r) from the axis, as shown in Fig. 3.30*c*.

Multiple Reflections. For a charge in the vicinity of the intersection of two conducting planes, as q in the region of AOB of Fig. 3.30d, there might be a temptation to use only one image in each plane, as 1 and 2 of Fig. 3.30d. Although $+q$ at Q and $-q$ at 1 alone would give constant potential as required along OA, and $+q$ at Q and $-q$ at 2 alone would give constant potential along OB, the three charges together would give

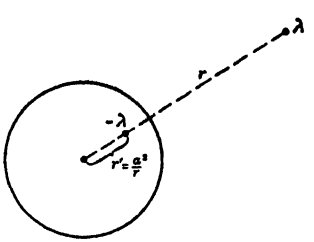


FIG. 3.30c. Image of a line charge in a conducting cylinder.

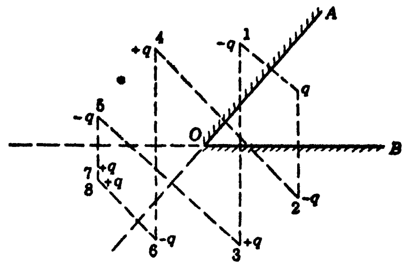


FIG. 3.30d. Multiple images in intersecting planes.

constant potential along neither OA nor OB. It is necessary to image these images in turn, repeating until further images coincide, or until all further images are too far distant from the region to influence potential. It is possible to satisfy exactly the required conditions with a finite number of images only if the angle AOB is an exact sub-multiple of 360°, as in the 45° case illustrated by Fig. 3.30d.

3.31 Example of the Use of Image Method

As an example of the application of the image theory, consider the following problem. In an electronic device, a beam of electrons is to pass between two conducting plates at the same potential. Thus there is no field in the space between the plates except that due to the electrons themselves. Now, the space charge repulsion between electrons in a beam is an important effect, often actually limiting the current that can be obtained in the beam. The question arises: is the tendency of the beam to spread due to mutual repulsions increased or decreased by directing it between the two conducting plates?

Suppose the beam is of circular cross section of diameter small compared with the distance between plates, very long compared with its diameter, and essentially uniform throughout its length in density and diameter. Of course, the extent to which the beam can be maintained in this condition depends, among other things, upon the answer to the question we are asking. Consider now Fig. 3.31 in which the beam is much closer to one plate than the other so that the effect of the distant

plate may be ignored. We shall consider two points on the edge of the beam, P and Q. First, it will be noted that the effect of the electrons in the beam is to cause an outward force, i.e., away from the beam axis.

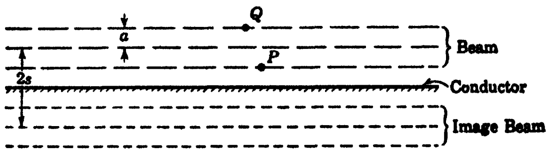


FIG. 3.31. Image of an electron beam in a conducting plane.

As a matter of fact, the outward field intensity at a radius r from the beam center is easily computed if we assume a long beam and use Gauss's law as in Art. 2.30.

$$E_1 = \frac{4\pi\rho}{2\pi r} = \frac{2\rho}{a}$$

where ρ is the charge density.

Now the image theory of Art. 3.30 tells us that the effect of the plate may be computed exactly by ignoring it and finding instead the field caused by an image beam of opposite sign charges. At point P there will be an attracting field due to the image beam of

$$E_2 = \frac{2\rho}{2s - a}$$

where $2s$ is the distance between the beam axis and its image axis. At point P, the net electric field which acts to spread out the electrons is

$$E_1 + E_2 = 2\rho\left[\frac{1}{a} + \frac{1}{2s - a}\right] = 2\rho\frac{2s}{a(2s - a)}$$

$$= E_1\frac{2s}{2s - a}$$

For $s = a$, where the beam just grazes the plate, the diverging electric field at the beam edge adjacent to the plate is twice as great as if the beam were far removed from any conductors.

It is interesting to note that at point Q there is actually a decrease in the force causing a divergence of the beam.

Problem 3.31. Repeat the example of Art. 3.31, making only the substitution of a semi-infinite dielectric slab for the plane conductor which is alongside the electron beam. This requires, of course, that you first discover how to set up images in the dielectric to give the electric field distribution in the space traversed by the beam.

4

MAXWELL'S EQUATIONS AND HIGH-FREQUENCY POTENTIAL CONCEPTS

THE LAWS OF TIME-VARIABLE ELECTRICAL PHENOMENA

4.01 Introduction

When the subject material of Chapter 2 was introduced, the objective was stated to be the derivation of a group of equations which would contain a description of fields due to static charges and static currents. It was claimed that in the solution of problems it would be well to have several forms for the statement of fundamental laws so that the most convenient might be chosen for the problem at hand. A number of the techniques involved in this process of selection of equations and their subsequent solution were discussed in Chapters 2 and 3.

In a similar way, an attempt will now be made to present the more complex theory that underlies electric and magnetic effects that vary with time. Of course, some of this theory is only an extension of the static theory. But the additional effects brought in by the varying of charges and currents very frequently complicate the solution of problems. Instead of presenting the different laws in equation form and giving methods of solution in one or two chapters, we shall require the rest of the book for the analysis of time-varying systems. In this chapter, a consistent set of equations describing varying electric and magnetic effects will be obtained. This material will serve as a basis for several subsequent chapters in which the equations will be applied to the field and wave problems of modern radio.

4.02 Voltages Induced by Changing Magnetic Fields

Faraday discovered experimentally that when the magnetic flux linking a closed circuit is altered, a voltage is induced in that circuit proportional to the rate of change of flux linking the circuit. This law is an experimental law of electricity and magnetism that requires little generalization to be widely useful. In the consideration of most circuits and electrical machinery it is necessary only to write:

$$V = n\frac{d\psi}{dt} \qquad [1]$$

where V is the voltage induced in a coil having n turns, and ψ is the flux linking the coil. The equation may be used directly to find the voltage induced by a generator coil moving in a magnetic field that varies with space, or to calculate the impedance presented by a coil to an alternating voltage. In the latter case the magnetic field due to the current in the coil may be calculated from the laws given in Chapter 2 if the frequency is not too high, and from this the flux linking the coil is found. Faraday's law applied to this flux permits the calculation of the voltage induced to oppose the applied voltage, hence the reactance drop. It is only when we are interested in phenomena that go far beyond these simple induction experiments, such as, for example, radio waves, that we find it necessary to generalize the law further.

Faraday's law gives a value for voltage induced in a circuit regardless of the resistance of that circuit, although, of course, the resulting current which flows will depend upon that resistance. It seems reasonable that this should be true even if the resistance of that path becomes infinite. In other words, the voltage around any closed path in space should be given by the rate of change of magnetic flux through that path. The voltage around the path is the line integral of voltage gradient (or negative line integral of electric intensity) so we may write this relation in terms of fields alone. Now, two sets of units have been defined: one applies to electric phenomena and the other to magnetic phenomena. The present situation, however, includes both phenomena. Before this chapter is over, we shall present all the equations in one consistent set of units which will be used throughout the text. For now, a constant c will be introduced to take care of the relation between units. Electric quantities will be written in electrostatic units and magnetic quantities in electromagnetic units just as they were introduced.

$$c\oint \bar{E} \cdot d\bar{l} = -\frac{d\psi}{dt}$$

Or, in terms of $\bar{H}$,

$$\oint \bar{E} \cdot d\bar{l} = -\frac{1}{c}\frac{\partial}{\partial t}\int_S \mu'\bar{H} \cdot d\bar{S} \qquad [2]$$

The partial derivative with time is used to distinguish it from variations in space, indicating that the law refers to the time rate of change of flux at a fixed region in space. From Stokes' theorem, Art. 2.26, and (2)

$$\int_S \nabla \times \bar{E} \cdot d\bar{S} = -\frac{1}{c}\frac{\partial}{\partial t}\int_S \mu'\bar{H} \cdot d\bar{S}$$

If this equation is to be true over any surface, no matter how small, the integrands must be equal.

$$\nabla \times \bar{E} = -\frac{\mu'}{c}\frac{\partial \bar{H}}{\partial t} \qquad [3]$$

The equation just obtained is a generalized differential expression of Faraday's law. The distinction should be noted here between this and the previous expressions of Chapter 2 for electric fields due to charges. There it was shown that the work integral is always zero.

$$\oint \bar{E}\cdot d\bar{l} = 0 \quad \text{or} \quad \nabla \times \bar{E} = 0$$

For electric fields arising from a change in magnetic field, the line integral about any path is no longer zero. The curl has a definite value given in terms of the time change in magnetic field. It may be recalled that the conservative property of the static electric field was derived from an energy consideration; hence it may seem that the existence of a curl for electric field shown above now violates the conservation of energy. This is fortunately not true. The existence of a curl tells only that the electric field is no longer conservative in itself if there is a changing magnetic field present, for there is then a transfer of energy from the magnetic to the electric field. Conservation of energy, in other words, tells us that since work is done if an electric charge is moved about a closed path in an electric field that comes about from a changing magnetic field, this work must have come out of magnetic field energy.

4.03 Continuity of Charge

Faraday's law is but one of the fundamental laws for changing fields. Certain other laws for electric and magnetic fields were derived in Chapter 2, but we should be somewhat suspicious of these, since they were determined by experiments on static systems. Let us assume for the moment that certain of these can be extended without revision to varying systems. Writing together the expressions for divergence and curl of electric and magnetic fields (differential expression for Gauss's law, Biot's law, Faraday's law)

$$\nabla \cdot \bar{D} = 4\pi\rho \qquad [1]$$

$$\nabla \cdot \bar{B} = 0 \qquad [2]$$

$$\nabla \times \bar{E} = -\frac{1}{c}\frac{\partial \bar{B}}{\partial t} \qquad [3]$$

$$\nabla \times \bar{H} = 4\pi\frac{\bar{\imath}}{c} \qquad [4]$$

$\bar{i}$ may be thought of as due to a motion of charges, say a charge density ρ moving with velocity $\bar{v}_\rho$. If ρ is expressed in electrostatic units and $\bar{H}$ in electromagnetic units, we must again introduce the conversion factor between units, c. In place of (4) we might then write

$$\nabla \times \bar{H} = 4\pi \left(\frac{\rho \bar{v}_\rho}{c} \right) \qquad [5]$$

With $i = \rho v_\rho$, and ρ in esu, i is in esu; hence the c in (4).

If these equations are to be applicable to systems which vary with time, they should satisfy certain requirements. For one thing, (1), (3), and (5) relate the three factors, $\bar{E}$, $\bar{H}$, and ρ. An elimination could be made to obtain an equation in ρ alone. We would expect this equation to show that however ρ may vary with space and time, it must vary in such a manner that total charge in the whole system is conserved. For any volume, if charge flows out, the amount inside must decrease. If, over the boundary surface, a net charge flows in, the total charge inside must increase. Considering a smaller and smaller volume, in the limit, the divergence of $\rho \bar{v}_\rho$, or the outward flow of charge per unit volume per unit time in space, must be the negative of the time rate of change of charge per unit volume at that point.

$$\nabla \cdot (\rho \bar{v}_\rho) = -\frac{\partial \rho}{\partial t} \qquad [6]$$

If, however, we take the divergence of $(\rho \bar{v}_\rho)$ from (5)

$$\nabla \cdot (\rho \bar{v}_\rho) = \frac{c}{4\pi} \nabla \cdot (\nabla \times \bar{H}) \equiv 0$$

which is not at all what was expected, and seems to indicate that (5), borrowed from statics, is not complete. Suppose that to (5) is added an additional vector term, $\bar{F}$.

$$\nabla \times \bar{H} = 4\pi \frac{\rho \bar{v}_\rho}{c} + \bar{F}$$

Then

$$\nabla \cdot (\rho \bar{v}_\rho) = -\frac{c}{4\pi} \nabla \cdot \bar{F}$$

If the continuity equation is to be satisfied

$$\frac{c}{4\pi} \nabla \cdot \bar{F} = \frac{\partial \rho}{\partial t}$$

But from (1)

$$\frac{\partial \rho}{\partial t} = \frac{1}{4\pi} \nabla \cdot \frac{\partial \bar{D}}{\partial t}$$

So the added factor must be

$$\bar{F} = \frac{1}{c}\frac{\partial \bar{D}}{\partial t}$$

Thus, finally

$$\nabla \times \bar{H} = 4\pi\left(\frac{\rho \bar{v}_\rho}{c}\right) + \frac{1}{c}\frac{\partial \bar{D}}{\partial t} \qquad [7]$$

is an equation which satisfies the requirements for continuity of charge.

4.04 The Concept of Displacement Current

In so far as the line integral of magnetic field is concerned, the term added to Eq. 4.03(5) might be considered as another current. Since it arises from the displacement vector $\bar{D}$, it may be called a displacement current as distinguished from the familiar current of charge motion, conduction or convection current of Eqs. 4.03(4) and 4.03(5). That is,

$$\nabla \times \bar{H} = \frac{4\pi}{c}(\bar{\imath}_c + \bar{\imath}_d)$$

where

$$\bar{\imath}_c = \text{convection current density} = \rho\bar{v}_\rho.$$

$$\bar{\imath}_d = \text{displacement current density} = \frac{1}{4\pi}\frac{\partial \bar{D}}{\partial t}.$$

In the next articles we shall try to see something of the physical significance of this new current. For the present let us ask why this current was not discovered earlier and why it has not been as common to us as convection current.

If representative calculations are made, it is apparent why the new term was not discovered in the experimental measurements checking Ampère's law at low frequencies, for its magnitude is very small until high frequencies are reached. For example, at a frequency of 10^6 cycles per second, field strength of 100 volts per centimeter, dielectric constant 5, the displacement current has a value of 2.80×10^{-4} ampere per square centimeter. Of course, at zero frequency (static case) the displacement current disappears altogether, but there is no violation of the continuity equation, since in this special case

$$\frac{\partial \rho}{\partial t} = 0$$

and consequently,

$$\nabla \cdot (\rho \bar{v}_\rho) = 0$$

4.05 Displacement Current in a Condenser

Now that the displacement current term has been acquired, we should be much happier about the problem of varying fields, for it is now possible to explain certain other things that should have proved worrisome had only conduction current been included in the law of Biot and Savart. Consider, for example, the circuit including the A-C generator and the condenser of Fig. 4.05. Suppose it is required to evaluate the line integral of magnetic field around the loop *a-b-c-d-a*. The law from statics states that the result obtained should be 4π times the current enclosed, that is, the current through any surface of which the loop is a boundary. Then it is true that if we take as the arbitrary surface through which current is to be evaluated one which cuts the wire A, as S_1, a finite value is obtained for $\oint \bar{H} \cdot d\bar{l}$. But suppose the surface selected is one which does not cut the wire, but instead passes between the plates of the condenser, as S_2. If conduction current alone were included, the computation would have indicated no current passing through this surface and the result would be zero. The path around which the integral is evaluated is the same in each case, and it would be quite annoying to possess two different results. It is the displacement current term which appears at this point to preserve the continuity of current between the plates of the condenser, giving the same answer in either case.

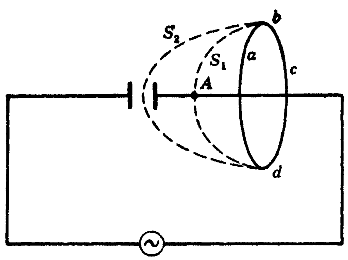


FIG. 4.05. Evaluation of $\oint \bar{H} \cdot d\bar{l}$ for an A–C circuit with a condenser.

To show that this continuity is preserved, consider a parallel plate condenser of capacity C, spacing d, area of plates A and applied voltage $V_0 \sin \omega t$. From circuit theory the charging current

$$I_c = C\frac{dV}{dt} = \omega C V_0 \cos \omega t$$

The field inside the condenser has a magnitude $E = V/d$ so the displacement current density is

$$i_d = \frac{\epsilon'}{4\pi}\frac{\partial \bar{E}}{\partial t}$$

$$i_d = \frac{\epsilon'}{4\pi d} V_0 \omega \cos \omega t$$

Total displacement current flowing between the plates is the area of the plate multiplied by the density of displacement current.

$$I_d = Ai_d = \omega\left(\frac{\epsilon' A}{4\pi d}\right) V_0 \cos \omega t$$

$$= \omega C V_0 \cos \omega t$$

This value for total displacement current flowing between the condenser plates is then exactly the same as the value of charging current flowing in the leads, calculated by the usual circuit methods above.

4.06 Displacement Current Due to a Moving Charge

Inclusion of the displacement current term is necessary for a valid discussion of another familiar case, that of a point charge in motion (Fig. 4.06). Suppose the charge q is moving with velocity $\bar{v}$ toward the mathematical loop A around which a conscientious observer is engaged in measuring the line integral, $\oint \bar{H} \cdot d\bar{l}$. If he has been given the equation

$$\oint \bar{H} \cdot d\bar{l} = 4\pi I$$

but has never been informed about displacement current, he is likely to be very unhappy. He has been told that he may take any one of the infinite number of possible surfaces of which his loop is the boundary, and after determining current passing through that surface, should then equate $\oint \bar{H} \cdot d\bar{l}$ to 4π times that current. Certainly at any instant there seem to be a large number of surfaces, such as S_1, through which there is no convection current (no motion of charge). Thus he feels that the integral must have a value of zero. But suppose he were to select the one surface, S_2, through which the charge q is passing at that instant. There would certainly be a current through this surface and a consequent value of $\oint \bar{H} \cdot d\bar{l}$ other than zero at that instant.

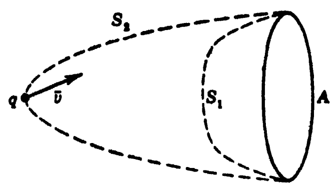


FIG. 4.06. Evaluation of $\oint \bar{H} \cdot d\bar{l}$ for charge q moving toward loop A.

This contradiction is again eliminated if the displacement current term is included. The line integral will then turn out to be the same no matter what surface is selected. In more general cases the contribution

to $\oint \bar{H} \cdot d\bar{l}$ may be part convection and part displacement currents for any given surface, and a different fraction convection and displacement currents for another surface, but the final result is always the same regardless of the surface chosen.

4.07 Maxwell's Equations

We have now the four general equations of electricity and magnetism:

$$\nabla \cdot \bar{D} = 4\pi\rho$$

$$\nabla \cdot \bar{B} = 0$$

$$\nabla \times \bar{E} = -\frac{1}{c}\frac{\partial \bar{B}}{\partial t}$$

$$\nabla \times \bar{H} = \frac{1}{c}\left(4\pi\bar{\imath} + \frac{\partial \bar{D}}{\partial t}\right)$$

These were first derived by Maxwell and are consequently known as Maxwell's equations. They are a complete statement of the relation between electric and magnetic fields and the currents and charges which give rise to them.

Specifically the equations give the relations *at any point* in a medium between charge density, conduction current density, and the electric and magnetic fields *at that point* due to the space distribution of currents and charges. The equations hold for any point in a medium, whether it is conducting or non-conducting, whether a vacuum or some dielectric or magnetic material. The $\bar{E}$ and $\bar{H}$ which appear in the equations are the total electric and magnetic fields, not just the contributions from the charges and currents at that point. Integral expressions will be obtained later which will permit us to calculate $\bar{E}$ and $\bar{H}$ from a knowledge of the charges and currents of the system. The practical problems utilizing Maxwell's equations in following chapters will clarify the theory contained in these equations.

A CLARIFICATION AND A CHOICE OF UNITS

4.08 Gaussian Units

Units used so far have been merely those most convenient. In Chapter 2 static electric effects and static magnetic effects were considered separately, so the most convenient separate units (esu and emu respectively) were used. In Chapter 3 the choice of units made little difference since the problem was essentially one of distribution, and one could hardly go wrong. So far in this chapter we have been interested

primarily in presenting the concepts of time-varying systems and so have again used the most convenient combination of units: esu for electric quantities, emu for magnetic quantities. Now that the final set of equations has been obtained, all the results should be stated in a consistent system of units convenient for use throughout the remainder of the book. The system of units chosen is the mks system of practical units. The other systems commonly used in engineering and physics texts will first be summarized; the reason for the units selected will be discussed; finally conversions between units will be given so that all references may be used without great difficulty.

If in Maxwell's equations all electric quantities are expressed in esu and all magnetic quantities in emu, it is then necessary to introduce a conversion factor between units, c, as was done. The system of units resulting is known as the Gaussian system of units. The equations were written in this system in Art. 4.07. Here $\bar{D}$, $\bar{E}$, ρ, and $\bar{\imath}$ (to be consistent with $\bar{\imath} = \rho\bar{v}_\rho$) were expressed in esu; $\bar{B}$ and $\bar{H}$ were expressed in emu. The conversion factor between units, c, has the dimensions of velocity and a magnitude which turns out to be the velocity of light in free space, as we shall see later.

4.09 Heaviside-Lorentz Rational Units

For extensive use of the foregoing equations, the factor 4π may become a nuisance. It is then convenient to define a new system of units which eliminates the factor 4π from the equations. From the unit charge in such a system there must then issue only one line of flux instead of the 4π lines found previously for the esu unit charge. Any system in which unit charges are chosen so that they are responsible for only one line of flux is known as a rational system of units, since it eliminates the irrational number 4π. The first such rational system was used by Heaviside and Lorentz, so is called the Heaviside-Lorentz system of units (abbreviated hlu). It is based upon the Gaussian system and so retains the factor $1/c$ in the equations. With all quantities in hlu,

$$\nabla \cdot \bar{D} = \rho \qquad \nabla \times \bar{E} = -\frac{1}{c}\frac{\partial \bar{B}}{\partial t}$$

$$\nabla \cdot \bar{B} = 0 \qquad \nabla \times \bar{H} = \frac{1}{c}\left(\bar{\imath} + \frac{\partial \bar{D}}{\partial t}\right)$$

4.10 Electrostatic and Electromagnetic Units

It is evident that all quantities might be written in either esu or emu. By using the conversion factor c properly, the two forms of the equations

are found to be

$$\text{esu:}\quad \nabla\cdot\bar{D} = 4\pi\rho \qquad \nabla\times\bar{E} = -\frac{1}{c^2}\frac{\partial\bar{B}}{\partial t}$$

$$\nabla\cdot\bar{B} = 0 \qquad \nabla\times\bar{H} = 4\pi\bar{\imath} + \frac{\partial\bar{D}}{\partial t}$$

$$\text{emu:}\quad \nabla\cdot\bar{D} = 4\pi c^2\rho \qquad \nabla\times\bar{E} = -\frac{\partial\bar{B}}{\partial t}$$

$$\nabla\cdot\bar{B} = 0 \qquad \nabla\times\bar{H} = 4\pi\bar{\imath} + \frac{1}{c^2}\frac{\partial\bar{D}}{\partial t}$$

4.11 Mks Practical Units

In all the previously described systems of units,

$$\bar{B} = \mu'\bar{H}$$

and

$$\bar{D} = \epsilon'\bar{E}$$

μ' and ϵ' are unity for free space. A study of the sets of equations for emu and esu shows that if the requirement that permeability and dielectric constant be unity for free space be removed, all constants 4π and c^2 may be absorbed in them. Then the equations would appear in the simplest possible form. We shall perform this simplification, first placing all quantities in a practical system of units that will be convenient for use in engineering, as follows:

- $\bar{E}$ electric intensity in volts/meter.
- ρ charge density in coulombs/meter^3.
- $\bar{\imath}$ current density in amperes/meter^2.
- $\bar{H}$ magnetic intensity in ampere turns/meter (or simply amperes/meter)
- $\bar{B}$ magnetic flux density in webers/meter^2 (a changing flux of 1 weber/second generates 1 volt)

All lengths are measured in meters.

This set of units is evidently based upon the mks (meter-kilogram-second) practical system. Using these units, the equations will be written with all constants absorbed in the permeability and dielectric constant, so that Maxwell's equations appear in their simplest possible form.

$$\nabla\cdot\bar{D} = \rho \tag{1}$$

$$\nabla\cdot\bar{B} = 0 \tag{2}$$

$$\nabla \times \bar{E} = -\frac{\partial \bar{B}}{\partial t} \quad [3]$$

$$\nabla \times \bar{H} = \bar{\imath} + \frac{\partial \bar{D}}{\partial t} \quad [4]$$

where $\bar{B} = \mu\bar{H}$ and $\bar{D} = \epsilon\bar{E}$.

4.12 Dielectric Constant and Permeability in Rational Mks Units

The use of rational mks units lumps all the inconveniences such as 4π's and conversion factors in ϵ and μ which must now be re-evaluated. Once the dimensions and magnitudes of μ and ϵ for various materials have been learned we shall then be prepared to use mks units throughout the remainder of the text. Since the units for all other quantities in Maxwell's equations have been decided upon, and since it is known that μ and ϵ must have such values as to maintain the accuracy of the equations, any information about μ and ϵ can best be obtained by an examination of those equations.

Setting $\bar{D} = \epsilon\bar{E}$ in the first of these equations, and recognizing that the divergence represents a differentiation with respect to distance, we see that

$$\frac{\epsilon(\text{Volts/meter})}{\text{Meter}} = \frac{\text{Coulombs}}{\text{Meter}^3}$$

or

$$\epsilon = \frac{\text{Coulombs}}{(\text{Volts})(\text{Meters})} = \frac{\text{Farads}}{\text{Meter}}$$

The dimensions of ϵ are known; its magnitude may next be found by converting the above quantities to esu for which we already know the relation between electric intensity and charge density for free space. Thus,

$$\epsilon = \frac{(\frac{1}{3} \times 10^{-9}\ \text{Statcoulomb})}{(300 \times \text{Statvolts})(\frac{1}{100} \times \text{Centimeters})}$$

$$= \frac{1}{9} \times 10^{-9} \left[\frac{\text{Statcoulomb/centimeter}^3}{\text{Statvolt/centimeter}^2}\right] \quad [1]$$

Now, the quantity in the brackets may be identified as $\rho/\nabla \cdot \bar{E}$ which for free space, in esu, of course, is $1/4\pi$. Then, if ϵ_0 denotes ϵ for free space ($\epsilon = \epsilon'\epsilon_0$)

$$\epsilon_0 = \frac{1}{36\pi} \times 10^{-9} \quad \text{farad/meter} \quad [2]$$

By similar steps and by using other Maxwell equations,

$$\mu_0 = 4\pi \times 10^{-7} \quad \text{henry/meter} \tag{3}$$

A system of units has now been determined in which Maxwell's equations are in the simplest possible form and in which current is in amperes, electric fields are in volts per meter, and magnetic fields are in ampere turns per meter. Impedances will be in ohms and power in watts. The only real disadvantage is that we must remember that μ and ϵ are not unity in vacuum. This is only a small disadvantage since it is much easier to remember the two values of μ_0 and ϵ_0 than the many conversion factors between practical, esu, and emu systems of units. Moreover, we shall find later that most often μ and ϵ appear in the easily remembered combinations $1/\sqrt{\mu\epsilon}$ and $\sqrt{\mu/\epsilon}$. These are already familiar since $1/\sqrt{\mu_0\epsilon_0}$ is the velocity of light in free space, 3×10^8 meters/second; $\sqrt{\mu_0/\epsilon_0}$ has dimensions of ohms and is 120π or about 377 ohms (by coincidence, the well-known ω for 60 cycles).

Problem 4.12(*a*). By using the divergence theorem show that the bracketed expression in Eq. 4.12(1) is indeed equal to $1/4\pi$ as stated.

Problem 4.12(*b*). Derive Eq. 4.12(3).

4.13 Cgs Practical Units

The obvious engineering advantages of the mks system of units might equally as well have been applied to the equations in cgs (centimeter-gram-second) practical units. All equations arrived at in this book may be used to obtain answers in cgs units if desired, by using the following values of μ_0, ϵ_0, and c in place of the values given in Art. 4.12 for mks units.

$$\epsilon_0 = \frac{1}{36\pi} \times 10^{-11} \text{ farad/centimeter.}$$

$$\mu_0 = 4\pi \times 10^{-9} \text{ henry/centimeter.}$$

$$\frac{1}{\sqrt{\mu_0\epsilon_0}} = c = 3 \times 10^{10} \text{ centimeters/second.}$$

$$\sqrt{\frac{\mu_0}{\epsilon_0}} = 120\pi \cong 377 \text{ ohms.}$$

In Maxwell's equations, the units would then be:

$\bar{E}$ volts/centimeter.
ρ coulombs/centimeter^3.
$\bar{i}$ amperes/centimeter^2.
$\bar{H}$ ampere turns/centimeter.
$\bar{B}$ webers/centimeter^2.

4.14 Units Conversion Table

So that all references may be used conveniently, conversions between all the above systems of units are tabulated below. It is then necessary only to multiply any results in mks rationalized units in this book by the factors shown to obtain results in any other desired system of units.

Multiply	by	To Obtain
1. Coulombs	$\frac{1}{10}$	abcoulombs
Coulombs	3×10^9	statcoulombs
Coulombs	$\dfrac{3 \times 10^9}{\sqrt{4\pi}} = 8.46 \times 10^8$	hlucoulombs
2. Amperes	$\frac{1}{10}$	abamperes
Amperes	3×10^9	statamperes
Amperes	$\dfrac{3 \times 10^9}{\sqrt{4\pi}} = 8.46 \times 10^8$	hluamperes
3. Volts	10^8	abvolts
Volts	$\frac{1}{300}$	statvolts
Volts	$\dfrac{\sqrt{4\pi}}{300} = 0.0118$	hluvolts
4. Ohms	10^9	abohms
Ohms	$\frac{1}{9} \times 10^{-11}$	statohms
Ohms	$\dfrac{4\pi}{9} \times 10^{-11}$	hluohms
5. Farads	9×10^{11}	statfarads
6. Henrys	10^9	abhenrys
7. Watts (joules/second)	10^7	ergs/second
8. Volts/meter	$\frac{1}{3} \times 10^{-4}$	statvolt/centimeter
9. Webers	10^8	maxwells
10. Webers/meter^2	10^4	gauss

POTENTIALS USED WITH VARYING CHARGES AND CURRENTS

4.15 Inadequacy of Scalar Potentials for Non-Static Electric Fields

The set of differential equations known as Maxwell's equations, with certain auxiliary relations, gives the complete information for obtaining electric and magnetic effects due to currents and charges. It will sometimes be convenient to put the information in a different form by the introduction of new variables. In the study of static fields, it was found that new functions known as potentials helped in the solution of static problems. We might then look for similar potential functions which will help in the solution of more general problems. The potential functions of static fields were given in terms of integral expressions of charges and

currents; the more general potential functions will be given in terms of different integral expressions of charges and currents. The static fields were obtained from the static potentials by differentiation; time-varying fields will be obtained from the new potentials by similar differentiations. All new expressions obtained will reduce to the original static expressions as time derivatives become zero.

The static electric field was derivable as the gradient of a scalar potential. When magnetic fields were encountered, it was not always permissible to use a scalar potential, since the line integral of magnetic field intensity about a closed path was, in general, not zero; that is, the field possessed a finite curl. In time-varying effects, the curl of electric field intensity is also not zero, but is rather given by Eq. 4.11(3) and the electric field can no longer be derived as the gradient of a scalar potential.

There have been previous discussions of potential functions which are both scalars and vectors. Consequently, let us speculate somewhat on the type of potential functions that might be used for electric intensity. The magnetic field in the static case was derivable as the curl of a vector potential, $\bar{A}$. However, if an attempt were made to obtain $\bar{E}$ from a similar vector function, say by setting

$$\bar{E} = \nabla \times \bar{F}$$

it would be found at once that $\nabla \cdot \bar{E}$ would be everywhere zero since div curl $\bar{F} \equiv 0$. The divergence should not be zero but should have the value given by Eq. 4.11(1). Thus we are faced with this problem: the electric intensity for time-varying fields cannot be derived alone as the gradient of a scalar potential since this would require that it have zero curl, and it actually has a finite curl of value $-\partial \bar{B}/\partial t$; it cannot be derived alone as the curl of a vector potential, since this would require that it have zero divergence, and it actually has a finite divergence of value ρ/ϵ.

Since the divergence of magnetic field is zero in the general case as it was in the static, it seems that $\bar{H}$ may still be set equal to the curl of some magnetic vector potential, $\bar{A}$. Suppose the substitution of $\bar{H} = \nabla \times \bar{A}$ is made in Maxwell's equations and an attempt is then made to obtain a value for the potential function of electric fields which vary with time. For a homogeneous medium in which μ and ϵ are constant, Eq. 4.11(3) becomes

$$\nabla \times \bar{E} + \mu \frac{\partial}{\partial t} (\nabla \times \bar{A}) = 0$$

or

$$\nabla \times \left[\bar{E} + \mu \frac{\partial \bar{A}}{\partial t} \right] = 0 \tag{1}$$

This equation states that the curl of a certain vector quantity is zero. But this is the condition that permits a vector to be derived as the gradient of a scalar, say Φ. That is,

$$\bar{E} + \mu \frac{\partial \bar{A}}{\partial t} = -\nabla\Phi$$

or

$$\bar{E} = -\nabla\Phi - \mu \frac{\partial \bar{A}}{\partial t} \tag{2}$$

The electric field, $\bar{E}$, has consequently been obtained in terms of both a scalar and a vector potential.

To continue the substitutions in Maxwell's equations: in Eq. 4.11(1)

$$-\nabla^2\Phi - \mu \frac{\partial}{\partial t} (\nabla \cdot \bar{A}) = \frac{\rho}{\epsilon} \tag{3}$$

in Eq. 4.11(4)

$$\nabla \times \nabla \times \bar{A} = \bar{\imath} + \epsilon \left[-\nabla \left(\frac{\partial \Phi}{\partial t} \right) - \mu \frac{\partial^2 \bar{A}}{\partial t^2} \right]$$

The vector identity, (Art. 2.38)

$$\nabla \times \nabla \times \bar{A} = \nabla(\nabla \cdot \bar{A}) - \nabla^2 \bar{A}$$

so

$$\nabla(\nabla \cdot \bar{A}) - \nabla^2 \bar{A} = \bar{\imath} - \epsilon \nabla \left(\frac{\partial \Phi}{\partial t} \right) - \mu\epsilon \frac{\partial^2 \bar{A}}{\partial t^2} \tag{4}$$

The two equations, (3) and (4), scarcely seem simple. Recalling the argument of Art. 2.29, it is realized that $\bar{A}$ is not unique until it is further specified. That is, there are any number of vector functions whose curl is the same. It may be shown that it is necessary only to specify the divergence of $\bar{A}$ to make it unique, and this may be done according to convenience. If the divergence of $\bar{A}$ is chosen as

$$\nabla \cdot \bar{A} = -\epsilon \frac{\partial \Phi}{\partial t}$$

(3) and (4) then simplify to

$$\nabla^2 \Phi - \mu\epsilon \frac{\partial^2 \Phi}{\partial t^2} = -\frac{\rho}{\epsilon} \tag{5}$$

$$\nabla^2 \bar{A} - \mu\epsilon \frac{\partial^2 \bar{A}}{\partial t^2} = -\bar{\imath} \tag{6}$$

A check to see if these potential functions and equations reduce to the familiar static expressions when time derivatives become zero shows that $\bar{H} = \nabla \times \bar{A}$, $\bar{E} = -\nabla\Phi$ as before, the condition for uniquely specifying $\bar{A}$ reduces to $\nabla \cdot \bar{A} = 0$, and (5) and (6) reduce to the two Poisson equations,

$$\nabla^2\Phi = -\frac{\rho}{\epsilon} \qquad [7]$$

$$\nabla^2\bar{A} = -\bar{\imath} \qquad [8]$$

So Φ and $\bar{A}$ in this special case are seen to be the same potentials selected in Chapter 2 except for the camouflage of different units.

Problem 4.15. Explain why it is necessary only to specify the divergence of $\bar{A}$ to make it unique. In other words, if $\bar{A}$ is to be a vector whose curl is $\bar{H}$ and whose divergence is ψ, it is the only such vector.

4.16 The Retarded Potentials

Equations 4.15(7) and 4.15(8) can be interpreted as saying that Φ is due to charges and $\bar{A}$ to currents. Thus $\bar{E}$ must be expressed in terms of both Φ and $\bar{A}$ since electric field may be due either to charges or to changes in the magnetic field. The term in Φ includes the effect of charges; the term in $\bar{A}$ includes the effect of changing magnetic fields as expressed by Faraday's law. This may seem rather matter-of-fact and obvious, but it cannot be overemphasized, for it is easy to fall into the error of thinking of time-varying effects in terms of a scalar potential applied between two points, with electric field given as the negative gradient of this potential as in the static case. This must often lead to incorrect conclusions, as shown by the general need for two types of potentials in the equation for $\bar{E}$.

The solutions of the two equations will be discussed later, but it has been shown that the solutions must reduce to the familiar integral expressions for potentials of the static case when time variations disappear. The additional terms in the differential equations are time derivatives, and so it would be expected that the solutions will differ from those of the static case in some fashion resulting from these time terms. This results in solutions in integral form quite similar to those of the static case, with the exception that the time term appears as a retardation effect. That is, it requires some time for an effect to be felt at a point if a charge or current is suddenly changed at another point. The potentials are accordingly called the *retarded potentials*.

4.17 Solution of the Potential Equations

To consolidate our gains, we have now the equations

$$\nabla^2 \Phi - \mu\epsilon \frac{\partial^2 \Phi}{\partial t^2} = -\frac{\rho}{\epsilon} \tag{1}$$

$$\nabla^2 \bar{A} - \mu\epsilon \frac{\partial^2 \bar{A}}{\partial t^2} = -\bar{\imath} \tag{2}$$

$$\bar{H} = \nabla \times \bar{A} \tag{3}$$

$$\bar{E} = -\nabla\Phi - \mu \frac{\partial \bar{A}}{\partial t} \tag{4}$$

These may be considered as expressions equivalent to Maxwell's equations, but they have the advantage that equations (1) and (2) are standard forms of differential equations. Had we attempted to eliminate directly in Maxwell's equations to obtain similar expressions containing $\bar{H}$ alone and $\bar{E}$ alone in terms of charges and currents, we would have found no such simple form as this without some such change of variable as that which brought in the new variables, $\bar{A}$ and Φ.

It is necessary to solve the two equations in $\bar{A}$ and Φ. The special case of no variations with time has been considered in Chapter 3. The special case of $\rho = 0$ and $\bar{\imath} = 0$ will be especially important since it covers radio waves propagating in free space. Later chapters will be devoted entirely to this subject. It is possible to obtain solutions to the above equations in integral form. Such solutions will be more general than those for any of the special cases mentioned and are helpful in getting a good picture of varying electrical effects. Although a rigorous solution of the equations is possible, it seems better suited to our present objectives merely to present the solutions to the equations, making them seem reasonable from our knowledge of similar equations of past chapters. The mathematical proof can be found in several of the reference books given in Appendix *A*.

Consider first (1). If we start with small time variations, or at least small compared with space variations, the equation will reduce to Poisson's equation of Chapter 2.

$$\nabla^2 \Phi = -\frac{\rho}{\epsilon}$$

The solution was found to be

$$\Phi = \frac{1}{\epsilon} \int_V \frac{\rho dV}{4\pi r}$$

Consequently, the solution should reduce to this form as r approaches zero.

On the other hand, in a region where there are no charges, $\rho = 0$, the equation becomes

$$\nabla^2 \Phi - \mu\epsilon \frac{\partial^2 \Phi}{\partial t^2} = 0$$

This is the wave equation, of which we have already considered the special case involving one space derivative only (Chapter 1). The solution of the wave equation represents a wave traveling with velocity $v = 1/\sqrt{\mu\epsilon}$. Tying together these two bits of information, it seems that the contribution to Φ at a point from a charge (ρdV) a distance r away should be of the same form as in the static case, with the exception that the effect requires a finite time, r/v, to propagate from the charge to the point; hence the contribution to Φ at time t should be calculated from the value of (ρdV) at time r/v before t, that is, at time $t - (r/v)$. Thus,

$$d\Phi = \frac{[\rho]_{t-\frac{r}{v}}\, dV}{4\pi\epsilon r}$$

$$\Phi = \int_V \frac{[\rho]_{t-\frac{r}{v}} dV}{4\pi\epsilon r} \qquad [5]$$

The bracket with subscript $t - (r/v)$ denotes that for an evaluation of Φ at time t, the value of ρ at time $t - (r/v)$ should be used in the integral.

By direct analogy between the equations in Φ and $\bar{A}$, the solution for $\bar{A}$ becomes

$$\bar{A} = \int_V \frac{[\bar{\imath}]_{t-\frac{r}{v}} dV}{4\pi r} \qquad [6]$$

This equation shows that $\bar{A}$ is also calculated by summing contributions from all convection currents of the system as before, except that the contribution to $\bar{A}$ at a given point at time t from a current element $(\bar{\imath} dV)$ a distance r away from that point should be calculated with the value of current $\bar{\imath}$ at time r/v before the time t; it required the time r/v to propagate its effect over that distance.

The velocity v is the velocity of light or electromagnetic waves in the homogeneous medium considered and is given in terms of the dielectric constant and permeability of that medium by

$$v = \frac{1}{\sqrt{\mu\epsilon}}$$

For free space

$$v = c = \frac{1}{\sqrt{\mu_0 \epsilon_0}} = 3 \times 10^8 \text{ meters/second}$$

In later articles the subscript $t - (r/v)$ will not always be written below the bracket. The use of brackets alone will be understood to indicate that the function inside is retarded; that is, it is a function of $t - (r/v)$ rather than t.

4.18 Electric and Magnetic Fields in Terms of Vector Potential Alone

Electric and magnetic fields have been expressed in terms of two potentials. One of these (Φ) is obtained in terms of all charges of the system. The other ($\bar{A}$) is obtained in terms of all currents of the system. Since by the continuity equation there is a relation between charges and currents, there must then be a corresponding relation between $\bar{A}$ and Φ. This has already been used in specifying the divergence of $\bar{A}$ (Art. 4.15)

$$\nabla \cdot \bar{A} = -\epsilon \frac{\partial \Phi}{\partial t} \tag{1}$$

With this relation between $\bar{A}$ and Φ, equations may be written for electric and magnetic potential in terms of one potential only. This will result in the most convenient form for use of the equations, although it will sometimes be desirable to retain both potentials to show more clearly the separate effects of charges and currents.

For steady state conditions, where all quantities vary as $e^{j\omega t}$, (1) becomes

$$\nabla \cdot \bar{A} = -j\omega\epsilon\Phi$$

or

$$\Phi = -\frac{1}{j\omega\epsilon}(\nabla \cdot \bar{A}) \tag{2}$$

The equation for electric field is then [Eq. 4.17(4)]

$$\bar{E} = -\nabla\Phi - j\omega\mu\bar{A} \tag{3}$$

or

$$\bar{E} = \frac{1}{j\omega\epsilon}\nabla(\nabla \cdot \bar{A}) - j\omega\mu\bar{A} \tag{4}$$

and, as before, $\bar{H} = \nabla \times \bar{A}$. Electric and magnetic fields are both expressed in terms of the one potential, $\bar{A}$, defined by Eq. 4.17(6).

BOUNDARY CONDITIONS FOR TIME-VARYING SYSTEMS

4.19 Matching Conditions for Tangential Electric Fields at a Boundary

In Chapter 2 there were obtained, as examples of the use of the laws of static fields, the matching conditions to be applied at the boundary between two materials. In succeeding chapters we shall be interested in the boundary conditions that must be applied for time-varying systems at the boundaries between two dielectrics, two conductors, or a conductor and a dielectric. These may be obtained directly from Maxwell's equations.

The integral form of Faraday's law from which

$$\nabla \times \bar{E} = -\frac{\partial \bar{B}}{\partial t}$$

was derived is

$$\oint \bar{E} \cdot d\bar{l} = -\int_S \frac{\partial \bar{B}}{\partial t} \cdot d\bar{S} \qquad [1]$$

Thus if at the boundary between any two materials an imaginary closed path is taken (Fig. 4.19) with its two sides an infinitesimal distance on either side of the boundary and Δl long, the line integral of electric field is

FIG. 4.19.

$$\oint \bar{E} \cdot d\bar{l} = (E_{t_1} - E_{t_2})\Delta l \qquad [2]$$

Since this path is an infinitesimal distance on either side of the boundary, it can include no area, and so can have no contribution from the surface integral of changing flux density. Consequently,

$$(E_{t_1} - E_{t_2})\,\Delta l = 0 \quad \text{or} \quad E_{t_1} = E_{t_2} \qquad [3]$$

This shows that tangential components of electric field are equal on the two sides of the boundary between any two materials.

4.20 Tangential Magnetic Fields at a Boundary

The integral form of

$$\nabla \times \bar{H} = \bar{\imath} + \frac{\partial \bar{D}}{\partial t}$$

is

$$\oint \bar{H} \cdot d\bar{l} = \int_S \left(\bar{\imath} + \frac{\partial \bar{D}}{\partial t}\right) \cdot d\bar{S} \qquad [1]$$

Just as for the development of Art. 4.19, an imaginary path taken as before will include no area, hence no contribution to the surface integral on the right, and so again

$$H_{t_1} = H_{t_2} \qquad [2]$$

This shows that tangential components of magnetic field are also equal on the two sides of the boundary between any two materials.

4.21 Tangential Magnetic Fields at a Current Sheet

It will be seen in later chapters that at high frequencies are found skin effect phenomena which effectively concentrate all current near the surface of conductors in a region negligibly small compared with the conductor dimensions. For these cases it is convenient to treat such currents effectively as current sheets of linear current density J amperes per meter. If the small closed path is taken only a small distance inside the conductor but still a distance deep enough into the conductors to include effectively all the current, there is a contribution to $\oint \bar{H} \cdot d\bar{l}$ from the conduction current encircled, although that from $\int_S \frac{\partial \bar{D}}{\partial t} \cdot d\bar{S}$ is still negligible. Then

$$(H_{t1} - H_{t2})dl = J_S dl \quad \text{or} \quad J_S = H_{t1} - H_{t2} \qquad [1]$$

J_S is the component of linear density of current flowing along the conducting boundary normal to the direction of $\bar{H}$.

This is exactly true only for a current sheet of infinitely small thickness.

4.22 Normal Components of Fields at a Boundary

The integral form of Gauss's law from which

$$\nabla \cdot \bar{D} = \rho$$

was derived is

$$\int_S \bar{D} \cdot d\bar{S} = \int_V \rho dV \qquad [1]$$

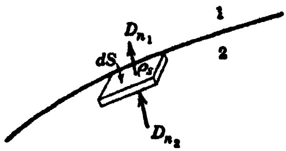


FIG. 4.22. Surface charge ρ_s on a boundary between two media.

If two very small elements of area ΔS are considered (Fig. 4.22), one on either side of the boundary between any two materials, with a surface charge density ρ_s sitting on the boundary, the application of Gauss's law to this elemental volume gives

$$(D_{n_1} - D_{n_2})dS = \rho_s dS$$

or

$$\rho_s = D_{n_1} - D_{n_2} \tag{2}$$

For a charge-free boundary,

$$D_{n_1} = D_{n_2} \quad \text{or} \quad \epsilon_1 E_{n_1} = \epsilon_2 E_{n_2}$$

and normal components of displacement are continuous.

Without isolated magnetic charges, a development exactly similar to that above shows that always

$$B_{n_1} = B_{n_2} \quad \text{or} \quad \mu_1 H_{n_1} = \mu_2 H_{n_2} \tag{3}$$

4.23 Use of the Boundary Relations for Time-Varying Problems

The previously discussed boundary relations are of the highest practical importance in the solution of high-frequency problems. They may enter into the problems in a number of ways. For example, the fields may be known on one side of a boundary, and the fields on the other side may be desired. Or the fields at the surface of a perfect conductor may be known, the current and charges on the conductor then being given by the boundary relations. But more important than these two examples is that the boundary relations are tied up basically with the whole technique of finding the distribution of electromagnetic effects by solving Maxwell's equations. In general, the problem is always one of writing down solutions to these equations and selecting or fitting them to the particular problem by making certain that they satisfy the boundary conditions of the space being studied. Hence the boundary relations appear directly or indirectly every time a high-frequency problem is solved.

The way in which these relations enter into the solution of a problem will be clarified by discussing a more or less general example. We will not actually solve this problem, but enough of the technique of solution will be gone into so that conclusions can be drawn about the application of boundary conditions to any problem. Figure 4.23 shows a space bounded everywhere, except at the boundary III, by a perfect conductor, I, and consisting of two regions of different dielectric materials A and B, whose boundary is denoted by II. It is assumed that the space described is excited by the impression of suitable boundary fields at III.

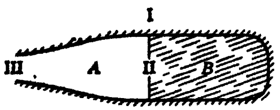


FIG. 4.23. Region containing two dielectrics partially enclosed by a conductor.

It will be assumed that only the time-varying parts of the phenomena are of interest here. To obtain the distribution of fields and currents and charges, various solutions of Maxwell's equations are now con-

sidered, and the process of selection of appropriate types and amounts of these solutions is ready to begin. In both regions A and B, we reject first of all those solutions which fail to give zero tangential electric field on boundary I. Other than this there are no further boundary conditions to apply to the solutions at this boundary. The proof of these statements is as follows. Since the conductor is perfect, there would be infinite current flow on its surface if there were any but zero tangential electric field there. Equation 4.19(3) states that if there is no tangential electric field at I in the dielectric, there will be no such field at I in the conductor. Thus the boundary condition of Eq. 4.19(3) is satisfied, and it has exerted its effect in selecting the proper solutions from all the candidate solutions. We next see if the other boundary conditions, Eqs. 4.21(1), 4.22(2), and 4.22(3), are automatically satisfied at I as implied above.

Equation 4.22(2) states that there must be a certain charge distribution on the conductor, if there is to be no normal electric field in the conductor. Equation 4.21(1) states that there must be a certain current flowing on the conductor, if there is to be no tangential magnetic field in the conductor. Since we know that all fields must vanish in the perfect conductor (time-varying effects only being considered), these two boundary conditions lead to a knowledge of the conductor's current and charge distribution. Notice that they normally are applied, in other words, *after* the solutions have been selected to fit the zero tangential electric field condition, and so do not take part in defining the solution. Of course, it is always possible that the whole problem might have been stated in reverse: Given a current distribution on I, what are the fields everywhere? In this problem, Eqs. 4.21(1) and 4.22(2) would obviously constitute constraints on the choice of solutions, but unless the current distribution were consistent with the requirements of zero tangential electric field on I, the problem proposed could never occur.

There remains at I the boundary condition expressed by Eq. 4.22(3). It is easily seen that this is automatically satisfied by the vanishing tangential electric field. The normal magnetic flux density $\bar{B}_n$ may be written, on the boundary, in terms of the component of the curl of the tangential electric field $\bar{E}_t$ on the boundary.

$$\nabla \times \bar{E}_t \Big|_{\text{boundary}} = -\frac{\partial \bar{B}_n}{\partial t} \qquad [1]$$

If $\bar{E}_t$ is everywhere zero on the boundary, then obviously $\bar{B}_n$ is likewise zero there and Eq. 4.22(3) becomes consistent with the vanishing of all fields in the conductor.

Next, boundary II must be considered. The solutions previously

selected must be tested and possibly rejected or determined so that the fields on the boundary II as obtained from expressions in region A or region B will be compatible; that is, the boundary conditions of Eqs. 4.22(2), 4.22(3), 4.20(2) and 4.19(3) must be met. Now, if the tangential field conditions are satisfied, it will be found that the normal field conditions are automatically taken care of. This may be shown in two ways. By writing the equation for the normal electric displacement vector on the boundary

$$\nabla \times \bar{H}_t \Big|_{\text{boundary}} = \frac{\partial \bar{D}_n}{\partial t} \qquad [2]$$

and considering this result together with (1), it is evident that specification of $\bar{E}_t$ and $\bar{H}_t$ fixes also $\bar{B}_n$ and $\bar{D}_n$ on the boundary. Another and perhaps more fundamental way of showing the interdependence of these boundary relations and the relation between tangential and normal fields is to recognize that the two equations

$$\nabla \cdot \bar{B} = 0 \quad \nabla \cdot \bar{D} = 0 \qquad [3]$$

may be derived from the two equations

$$\nabla \times \bar{H} = \frac{\partial \bar{D}}{\partial t} \quad \nabla \times \bar{E} = -\frac{\partial \bar{B}}{\partial t} \qquad [4]$$

We refer here, of course, only to the time-varying parts of the fields. This derivation is easily carried out and leads immediately to the conclusion that any relations, boundary or otherwise, based on (3) must be contained in (4).

That the specification of tangential fields is sufficient to determine the normal fields holds also, of course, at boundary III. Here the final application of the boundary conditions takes place. Of the solutions that have passed the previous boundary tests, the proper amount of each is selected to yield the given impressed tangential fields at III. Of course, it is evident from (1) and (2) that if normal fields are specified, some information is always obtained about the tangential fields, and often at a boundary it is convenient and possible to specify a normal field component in place of one of the tangential field components.

To summarize, for time-varying problems, the boundary conditions that will in general be used and which are sufficient for any such problem are:

At a perfectly conducting boundary, the tangential electric field must be set equal to zero.

At a boundary between two dielectric media, the tangential electric and magnetic fields must be made continuous.

Problem 4.23(a). Show that if in free space all phenomena are proportional to $e^{j\omega t}$, two of Maxwell's equations

$$\nabla \cdot \bar{B} = 0 \quad \nabla \cdot \bar{D} = 0$$

may be derived from the other two equations

$$\nabla \times \bar{H} = j\omega\bar{D} \quad \nabla \times \bar{E} = -j\omega\bar{B}$$

Show also that if the tangential fields are known at a boundary under the above conditions, the normal fields are also known.

Problem 4.23(b). Show that if continuity of charge is assumed,

$$\nabla \cdot \bar{B} = 0$$
$$\nabla \cdot \bar{D} = \rho$$

may be derived from

$$\nabla \times \bar{H} = \rho\bar{v}_\rho + \frac{\partial \bar{D}}{\partial t}$$

and

$$\nabla \times \bar{E} = -\frac{\partial \bar{B}}{\partial t}$$

This fact has made it quite common to include only the last two equations in referring to Maxwell's equations.

SUMMARY OF FIELD AND POTENTIAL EQUATIONS FOR TIME-VARYING SYSTEMS

4.24 General Equations and Definitions

We have now derived two separate forms for the laws of electric and magnetic systems with time variations: the differential equation form of Maxwell's equations and the statement in terms of retarded potentials. This information is all that is necessary to analyze all the following problems of circuits, circuit elements, antennas, resonators, wave guides, etc. It is consequently used so often that it is worth while to summarize the two statements of the laws with definitions of all quantities for easy reference.

Differential Equation Form	Retarded Potential Form
$\nabla \cdot \bar{D} = \rho$	$\bar{H} = \nabla \times \bar{A}$
$\nabla \cdot \bar{B} = 0$	$\bar{E} = -\nabla\Phi - \mu\dfrac{\partial \bar{A}}{\partial t}$
$\nabla \times \bar{E} = -\dfrac{\partial \bar{B}}{\partial t}$	$\bar{A} = \displaystyle\int_V \frac{[\bar{\imath}]_{t-\frac{r}{v}}\, dV}{4\pi r}$
$\nabla \times \bar{H} = \bar{\imath} + \dfrac{\partial \bar{D}}{\partial t}$	$\Phi = \displaystyle\int_V \frac{[\rho]_{t-\frac{r}{v}}\, dV}{4\pi\epsilon r}$
	$\nabla \cdot \bar{A} = -\epsilon\dfrac{\partial \Phi}{\partial t}$

	Definitions	Units
ρ	charge density	coulombs/meter^3
$\bar{\imath}$	current density In space charge regions, $\bar{\imath} = \rho\bar{v}_\rho$ In a conductor, $\bar{\imath} = \sigma\bar{E}$ (Ohm's law)	ampere/meter^2
$\bar{E}$	electric intensity (force equation $\bar{f} = q\bar{E}$)	volts/meter
$\bar{H}$	magnetic intensity (force equation $d\bar{f} = Id\bar{l} \times \bar{B}$)	amperes/meter
$\bar{B}$	magnetic flux density, $\bar{B} = \mu\bar{H}$	webers/meter^2
$\bar{D}$	electric flux density and displacement, $\bar{D} = \epsilon\bar{E}$	coulombs/meter^2
ϵ	dielectric constant	farads/meter
μ	permeability	henrys/meter
σ	conductivity	mhos/meter
v	velocity of light in the medium, $v = \dfrac{1}{\sqrt{\mu\epsilon}}$	meters/second
$\epsilon = \epsilon'\epsilon_0$	where ϵ_0 is the dielectric constant of free space, $\dfrac{1}{36\pi} \times 10^{-9}$	farad/meter
$\mu = \mu'\mu_0$	where μ_0 is the permeability of free space, $4\pi \times 10^{-7}$	henry/meter
$c = \dfrac{1}{\sqrt{\mu_0\epsilon_0}} = 3 \times 10^8$		meters/second

Stored Energy. Two expressions for energy in static fields which were derived in Chapter 2 (Arts. 2.33 and 2.34) will be found useful later. It will be convenient to have these expressions in mks units.

$$U_E = \tfrac{1}{2}\int_V \epsilon E^2\, dV \quad \text{joules}$$

$$U_H = \tfrac{1}{2}\int_V \mu H^2\, dV \quad \text{joules}$$

4.25 Steady State Alternating-Current Equations

For steady state A-C conditions in which all factors are assumed to vary as $e^{j\omega t}$, the field equations are:

Differential Equation Form	Retarded Potential Form
$\nabla \cdot \bar{D} = \rho$	$\bar{H} = \nabla \times \bar{A}$
$\nabla \cdot \bar{B} = 0$	$\bar{E} = \dfrac{1}{j\omega\epsilon}\nabla(\nabla \cdot \bar{A}) - j\omega\mu\bar{A}$
$\nabla \times \bar{E} = -j\omega\mu\bar{H}$	$\bar{A} = \displaystyle\int_V \frac{\bar{\imath}}{4\pi r} e^{-j\omega r/v}\, dV$
$\nabla \times \bar{H} = \bar{\imath} + j\omega\epsilon\bar{E}$	

4.26 Maxwell's Equations in Several Coordinate Systems

Rectangular Coordinates

1. $$\frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} + \frac{\partial D_z}{\partial z} = \rho$$

2. $$\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z} = 0$$

3. $$\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} = -\frac{\partial B_x}{\partial t}$$

$$\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} = -\frac{\partial B_y}{\partial t}$$

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = -\frac{\partial B_z}{\partial t}$$

4. $$\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} = \dot{i}_x + \frac{\partial D_x}{\partial t}$$

$$\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} = \dot{i}_y + \frac{\partial D_y}{\partial t}$$

$$\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = \dot{i}_z + \frac{\partial D_z}{\partial t}$$

Cylindrical Coordinates

1. $$\frac{1}{r}\frac{\partial}{\partial r}(rD_r) + \frac{1}{r}\frac{\partial D_\phi}{\partial \phi} + \frac{\partial D_z}{\partial z} = \rho$$

2. $$\frac{1}{r}\frac{\partial}{\partial r}(rB_r) + \frac{1}{r}\frac{\partial B_\phi}{\partial \phi} + \frac{\partial B_z}{\partial z} = 0$$

3. $$\frac{1}{r}\frac{\partial E_z}{\partial \phi} - \frac{\partial E_\phi}{\partial z} = -\frac{\partial B_r}{\partial t}$$

$$\frac{\partial E_r}{\partial z} - \frac{\partial E_z}{\partial r} = -\frac{\partial B_\phi}{\partial t}$$

$$\frac{1}{r}\frac{\partial}{\partial r}(rE_\phi) - \frac{1}{r}\frac{\partial E_r}{\partial \phi} = -\frac{\partial B_z}{\partial t}$$

4. $$\frac{1}{r}\frac{\partial H_z}{\partial \phi} - \frac{\partial H_\phi}{\partial z} = i_r + \frac{\partial D_r}{\partial t}$$

$$\frac{\partial H_r}{\partial z} - \frac{\partial H_z}{\partial r} = i_\phi + \frac{\partial D_\phi}{\partial t}$$

$$\frac{1}{r}\frac{\partial}{\partial r}(rH_\phi) - \frac{1}{r}\frac{\partial H_r}{\partial \phi} = i_z + \frac{\partial D_z}{\partial t}$$

Spherical Coordinates

1. $$\frac{1}{r^2}\frac{\partial}{\partial r}(r^2 D_r) + \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(D_\theta \sin\theta) + \frac{1}{r\sin\theta}\frac{\partial D_\phi}{\partial\phi} = \rho$$

2. $$\frac{1}{r^2}\frac{\partial}{\partial r}(r^2 B_r) + \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(B_\theta \sin\theta) + \frac{1}{r\sin\theta}\frac{\partial B_\phi}{\partial\phi} = 0$$

3. $$\frac{1}{r\sin\theta}\left[\frac{\partial}{\partial\theta}(E_\phi\sin\theta) - \frac{\partial E_\theta}{\partial\phi}\right] = -\frac{\partial B_r}{\partial t}$$

$$\frac{1}{r}\left[\frac{1}{\sin\theta}\frac{\partial E_r}{\partial\phi} - \frac{\partial}{\partial r}(rE_\phi)\right] = -\frac{\partial B_\theta}{\partial t}$$

$$\frac{1}{r}\left[\frac{\partial}{\partial r}(rE_\theta) - \frac{\partial E_r}{\partial\theta}\right] = -\frac{\partial B_\phi}{\partial t}$$

4. $$\frac{1}{r\sin\theta}\left[\frac{\partial}{\partial\theta}(H_\phi\sin\theta) - \frac{\partial H_\theta}{\partial\phi}\right] = i_r + \frac{\partial D_r}{\partial t}$$

$$\frac{1}{r}\left[\frac{1}{\sin\theta}\frac{\partial H_r}{\partial\phi} - \frac{\partial}{\partial r}(rH_\phi)\right] = i_\theta + \frac{\partial D_\theta}{\partial t}$$

$$\frac{1}{r}\left[\frac{\partial}{\partial r}(rH_\theta) - \frac{\partial H_r}{\partial\theta}\right] = i_\phi + \frac{\partial D_\phi}{\partial t}$$

5

CIRCUIT CONCEPTS AND THEIR VALIDITY AT HIGH FREQUENCY

THE FORMULATION OF A CIRCUIT CONCEPT CONSISTENT WITH MAXWELL'S EQUATIONS

5.01 Introduction

From the preceding chapter we have a set of laws, Maxwell's equations, that contains the core of the classical theory of electricity and magnetism. The applications of these to most problems is not difficult, speaking of concepts alone. There are plenty of mathematical difficulties — inability to integrate certain forms or to solve certain differential equations — but the ideas behind everything in modern radio, in so far as they depend on classical electricity and magnetism, should always be clarified by proper reference to Maxwell's equations. Our purpose for the remainder of the book is to study the systems and phenomena important to radio by means of these laws. They will be made to give quantitative design results, exact or approximate, whenever this is possible, but more important, we shall always use them to understand the concepts and physical pictures underlying the phenomena in question.

Of the many types of problems to be studied, many involve circuits, a term that covers a huge percentage of all phenomena with which the radio engineer is concerned and with which he associates many of the important concepts in electromagnetics. In a circuit problem there is often an applied voltage, and there are currents in the conductors of the circuit, charges on condensers in the circuit, ohmic losses, and power losses by radiation. These effects include almost everything that can happen when electric currents, charges, and conductors are let loose. The circuit problem is also one of the commonest problems illustrating the idea of cause and effect relationships. For these reasons, and, most important of all, because the circuit technique is one of the most familiar and useful to engineers, this will be the first problem to be investigated from the starting point of the fundamental laws.

In this chapter only the concepts and the general techniques of circuits are to be studied; quantitative analysis is reserved for the next chapter. From the rigorous starting point of the fundamental laws, it will be found that for circuits which are small compared with wavelength, this

exact approach leads directly to the familiar circuit ideas based upon Kirchhoff's laws, and the concepts of lumped inductances and capacitances are sufficient for analysis. For such circuits there would then be little need for going beyond Kirchhoff's laws. Although most of the circuits the average radio engineer encounters may be of this type, two reasons make it necessary to go beyond this stage in the understanding of circuit concepts. First, the increasing use of high frequencies increases the uncertainties in the engineering and development of systems which are thought up, designed, and experimented upon without sufficiently broad tools. For instance, most notions of circuits came out of studies of systems in which the current flows in relatively small cross-section filaments or wires (and in which the matter of distribution of current over this confined path is a secondary effect easily added on separately). But at ultra-high frequencies we would like to be able to use convenient circuit concepts, without going astray, on circuits which have the total current flow distributed over a wider or larger region than the physical confines of the circuit materials themselves. Secondly, when the radiation of electromagnetic energy is considered at any frequency, the radiating system must eventually be understood both as to the mechanism of the release of energy and the feeding of the antenna by the applied electromagnetic forces. The desire to utilize the convenient concepts of applied voltage, impedance, etc., in the latter case leads to a combination field-circuit problem, which, with no background in the electromagnetics of circuit notions, would be unnecessarily difficult even in qualitative thinking.

5.02 Applied Field and Resultant Current Density

Perhaps the most important single relation that appears in classical circuit theory is Ohm's law, which relates current flow to voltage drop in a conductor. This law may be generalized so that it applies to an infinitesimal conducting cube and is then written

$$\bar{\imath} = \sigma \bar{E} \qquad [1]$$

Equation (1) relates the current density at a point in a conductor to the electric field intensity at that point through the constant σ, known as the conductivity of the material. The electric intensity $\bar{E}$ is the total electric intensity at the point, not just a portion. The use of total field is emphasized now because the equations are to be put in a form suitable for exploitation by convenient, well-established circuit ideas. These include, near the top of the list, the concepts of applied voltage and voltage drops due to capacitive and inductive effects, as well as ohmic voltage drops. It should be recalled that in a circuit to which an

external voltage has been applied, the notions of circuit theory have us subtract from this applied voltage the back voltages or voltage drops due to the varying currents of the system and the varying charges of the system, leaving a certain net voltage available for the ohmic drop. To set the background for an approach to circuit ideas from the field equations, such a division will be followed in the electric fields and the notions of voltages will be arrived at by way of the fields.[1]

Thus, $\bar{E}$ may be made up of one part $\bar{E}_0$ applied from another system (the external generator) and another part $\bar{E}'$ arising from charges and currents in the circuit or system considered.

$$\bar{E} = \bar{E}_0 + \bar{E}' \qquad [2]$$

Recall that in Maxwell's equations, Art. 4.07, if all charges and currents are included in the equations, the electric intensity $\bar{E}$ appearing in the equations must be total electric intensity. If a system is considered which we have decided to call a circuit, and if this is influenced by another system which is the generator or source of applied voltage or applied field for the circuit, Maxwell's equations might, of course, be applied to the totality of the two systems, including all charges and currents for the circuit and its generator. Such an approach would be unnecessarily complicated if the generating system is, for all practical purposes, independent of the driven circuit. This is the case, for example, if the circuit obtains its applied field from an influencing system which is a distant antenna, a battery, a source of thermal emf, or a well-shielded signal generator. It is then easier to divide total field into two parts. There is the applied field which does not depend upon the charges and currents in the circuit, and there is an induced field which arises directly from these charges and currents. The basic laws applied only to the charges and currents of the circuit give only the amount of the induced field.

Total field, to be used in Ohm's law, is the sum of applied and induced components.

$$\frac{\bar{\imath}}{\sigma} = \bar{E}_0 + \bar{E}' \qquad [3]$$

The component $\bar{E}'$ due to charges and currents in the circuit may be stated conveniently in terms of the potentials (Art. 4.16):

$$\bar{E}' = -\nabla\Phi - \mu\frac{\partial \bar{A}}{\partial t}$$

[1] This follows closely the procedure of Carson, *Bell Syst. Tech. Jour.*, 6, 1–17 (January, 1927).

where Φ is the scalar potential due to charges of the system,

$$\Phi = \int_V \frac{[\rho]\,dV}{4\pi\epsilon r}$$

and $\bar{A}$ is the vector potential due to the currents of the system

$$\bar{A} = \int_V \frac{[\bar{\imath}]\,dV}{4\pi r}$$

Substituting these in (3),

$$\frac{\bar{\imath}}{\sigma} = \bar{E}_0 - \nabla\Phi - \mu\frac{\partial\bar{A}}{\partial t}$$

or

$$\bar{E}_0 = \frac{\bar{\imath}}{\sigma} + \nabla\Phi + \mu\frac{\partial\bar{A}}{\partial t} \qquad [4]$$

Equation (4) is the type of cause and effect relationship desired, since an applied field $\bar{E}_0$ results in an ohmic term and terms due to the charges and currents of the system. It is the first step in obtaining a circuit equation that relates voltages to currents and is based upon rigorous field theory. We must now define exactly what is meant by a circuit so that (4), which holds at a point in a conductor, may be extended to some proper integral relation that is true for a loop or circuit.

5.03 Applied Voltage and the Circuital Relations

A circuit will be defined only as a line in space. Of course, if there is to be any advantage in looking at the system as a circuit, this line or path will usually lie partially or entirely along a conductor. For any point, whether in a conductor or not, the relations between cause and effect are those derived from Maxwell's equations, Eq. 5.02(4). To obtain a circuit equation it is necessary only to integrate this differential expression along any path that is to be chosen as the circuit.

$$\int \bar{E}_0\cdot\overline{dl} = \int\frac{\bar{\imath}}{\sigma}\cdot\overline{dl} + \int\nabla\Phi\cdot\overline{dl} + \int\mu\frac{\partial\bar{A}}{\partial t}\cdot\overline{dl} \qquad [1]$$

The first term of the equation, $\int \bar{E}_0\cdot\overline{dl}$, is defined as the applied voltage of the circuit. More generally, this line integral of $\bar{E}_0$ between any two points a and b (Fig. 5.03) along the path of the circuit will be defined as the applied voltage of the circuit between the points a and b. Thus,

$$\int_a^b \bar{E}_0\cdot\overline{dl} \equiv V_{ba}$$

By voltage we shall mean no more than this. It should not be confused with the scalar potential of static fields.

The defined applied voltage of a circuit brings us to the first of several concepts which calls for careful handling to avoid confusion. First let us look at the easy case of direct current. Suppose it is made to flow in a conducting loop connected across the terminals of a battery. The battery voltage causes the current flow, and it is the only thing causing such a flow, since there is no electric field due to alternating currents, and there is no electric field in the conducting loop due to charges. (The capacitances of the battery plates and the conductor are neglected as being immaterial even if they are not small, since these capacitances, once charged, do not enter into current flow considerations.) Usual circuit theory would say that the battery applies a voltage between the two ends of the loop. Field theory says first that the battery must be applying an electric field in the conductor, otherwise there would be no current flow there. The two theories harmonize when applied voltage between two points is defined as the integral of applied electric field between those points. The circuit equations and concepts do not concern themselves with how the battery caused the voltage; neither do the field equations concern themselves with how it produced the applied field.

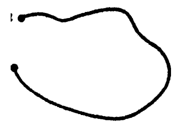


FIG. 5.03.

Consider next a closed loop of wire, the wire being of infinitesimally small cross-sectional area. Let magnetic flux through this loop be produced by some independent system which causes this flux to increase uniformly with time. The effect of this constant rate of change is to yield an applied D-C voltage, and by Ohm's law, this yields a certain direct current flow. If the field in the loop is oscillating in time, as in a receiving antenna excited by the field of a distant transmitting antenna, it is again clear that the applied voltage is the integral around the loop of the electric field due to the distant transmitter.

For the applied voltage produced by the battery, we did not know or care exactly how it was produced; it was known only that this voltage was of a certain amount and was independent of the path chosen for its circuit. However, when the applied voltage arises from the field of a distant antenna, the amount of this voltage depends very definitely upon the path of the circuit. It may be different for different sizes, orientations, and positions of the circuit. So, in general, the applied voltage around any loop to which the circuit concept is applied may vary radically in magnitude as different loops are selected, even when voltage is due to the same source.

With this notion of applied voltage, (1) is definitely suggestive of the Kirchhoff circuit equation for a simple series circuit containing resistance, inductance, and capacitance. The left-hand member is an applied voltage. This results in certain voltage drops expressed by the terms on the right. The first of these is an ohmic drop, the second is an induced drop from charges, and the third is an induced drop from changing magnetic effects. The relationship between these terms and the familiar resistance, capacitance, and inductance drops for low-frequency circuits will next be shown.

THE APPROXIMATE CIRCUIT EQUATION AND CIRCUIT CONSTANTS

5.04 Inductance of Circuits Small Compared with Wavelength

If Eq. 5.03(1) is applied to a closed circuit, there results the exact equation

$$V = \oint \frac{\bar{\imath}}{\sigma} \cdot d\bar{l} + \oint \mu \frac{\partial \bar{A}}{\partial t} \cdot d\bar{l} + \oint \nabla\Phi \cdot d\bar{l} \qquad [1]$$

It would be difficult to calculate numerical values from the equation in this form. Without making certain approximations there is little in (1) which shouts the advantages of using circuit concepts. Approximations applicable to typical circuits may often be:

1. Dimensions small compared with wavelength.
2. Current confined to a small filamentary conductor.
3. Current the same at all points about the path.

Let such a concentrated current, $\bar{I}$, be given by $\bar{a}_i I$. I is now the magnitude of current in amperes, and is the same everywhere about the path; $\bar{a}_i$ is a unit vector which gives the direction of current at any point. Further, we shall write a new constant that relates total current I (instead of current density $\bar{\imath}$) to total electric field. Call this total conductivity σ', and let its inverse be R'. For $\bar{\imath}/\sigma$, $\bar{a}_i I R'$ may be written, where R' is the resistance per unit length at any point. Since $\bar{a}_i$ and $d\bar{l}$ have the same direction at every point (that is direction of current at any point is that of its conducting path)

$$\oint \frac{\bar{\imath}}{\sigma} \cdot d\bar{l} = I \oint R' dl = RI$$

R is the total resistance about the path.

The line integral of $\nabla\Phi$ about any closed path must be zero, because Φ is a scalar potential (see Art. 2.14).

$$\oint \nabla\Phi \cdot d\bar{l} \equiv 0$$

There remains

$$V = IR + \frac{\partial}{\partial t}\oint \mu\bar{A} \cdot d\bar{l} \qquad [2]$$

The integral expression for $\bar{A}$, Eq. 4.17(6), for the case of current concentrated in a thin filament, becomes

$$\bar{A} = \oint \frac{[\bar{I}]dl'}{4\pi r} \qquad [3]$$

where r, now, is the distance between dl and dl'. If frequency is so low that the time necessary to propagate electromagnetic effects over any of the circuit dimensions is negligible compared with a period of the changing A-C effects (or in other words, the circuit is small compared with wavelength) retardation may be neglected and $\bar{A}$ written simply as

$$\bar{A} = I\oint \frac{\bar{a}_i dl'}{4\pi r}$$

I may be removed from the integral since it is assumed constant about the path, but $\bar{a}_i$ must, of course, be retained inside the integral since direction of current flow may change along the path. The evaluation of the integral then leads to some net directed quantity or vector which gives the direction of $\bar{A}$ and which is dependent only on circuit configuration, not upon magnitude of current.

$$\bar{A} \propto \bar{a}_A I$$

If a coefficient, L, is defined as

$$L \equiv \oint \frac{\mu\bar{A} \cdot d\bar{l}}{I} \qquad [4]$$

this will be a constant independent of current, since A by the above reasoning is proportional to current. This constant is a scalar and is dependent only upon the geometrical configuration and dimensions of the circuit.

Equation (2) may now be written

$$V = IR + L\frac{dI}{dt} \qquad [5]$$

(The time derivative can now be written as a total derivative since only time variations of current are being considered.) This is exactly the form of the low-frequency circuit equation for a series circuit with resistance and inductance. Let us examine further the defined coefficient L. From Stokes' theorem,

$$\oint \frac{\mu \bar{A} \cdot d\bar{l}}{I} = \int_S \frac{\mu(\nabla \times \bar{A}) \cdot d\bar{S}}{I}$$

But

$$\bar{H} = \nabla \times \bar{A} \quad \text{and} \quad \bar{B} = \mu \bar{H}$$

so

$$L = \frac{\int_S \bar{B} \cdot d\bar{S}}{I} \tag{6}$$

Since $\int_S \bar{B} \cdot d\bar{S}$ is the amount of magnetic flux passing through the circuit, (6) is the exact equivalent of the usual low-frequency definition, which defines inductance as the flux linkage per unit current, $L = \psi/I$.

It is, of course, not surprising that this result is obtained since the assumptions of this article are equivalent to those employed in more conventional derivations of this definition of inductance. However, the manner of deriving the equivalence here shows clearly the approximations, leads as we shall see later to new concepts of inductance, and is easily adapted to inclusion of new correction terms when the above approximations are not good approximations.

5.05 Capacitance Effects in Circuits Small Compared with Wavelength

The circuit of the preceding derivation was continuous Let us now break the circuit at some point. There is then the possibility of the accumulation of charge. The break we shall specify as rather small compared with other dimensions of the circuit, but plates may be placed at the discontinuity, if desired, to increase the possibility of accumulating charge. Thus, in spite of the camouflage of careful specification, a lumped capacitance has been inserted in an otherwise completely conducting filamentary path.

In Fig. 5.05 is shown the circuit with discontinuity. It might at first seem that the exact differential relation 5.02(4) could again be integrated around a closed path, the discontinuity ignored, and the term in $\nabla\Phi$ eliminated as before. However, if this is done, there is no way of know-

ing what happens to the term i/σ over the gap, for although current i is zero in the gap, so also is σ. This term is thus indeterminate. That it need not be zero is evident by recalling that it is the difference between applied and induced electric fields, and this difference can be made to have any value. In particular, consider the special case of all gap, that is, just an imaginary line in free space. Here, no matter what the applied field may be, there are no charges and no currents along the path, and so no induced field E' at all.

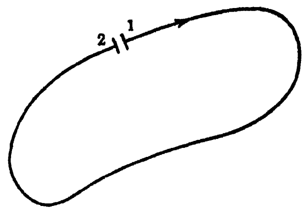


FIG. 5.05. Circuit containing a lumped capacitance.

If a circuit equation is to be written for a discontinuous conducting path, it is then to be obtained by integrating only over the conductor, from 1 to 2, where we know all terms of the equation. Integrating Eq. 5.02(4) from 1 to 2,

$$\int_1^2 \bar{E}_0 \cdot \bar{dl} = IR + \int_1^2 \mu \frac{\partial \bar{A}}{\partial t} \cdot \bar{dl} + \int_1^2 \frac{\partial \Phi}{\partial l}\, dl \qquad [1]$$

$$V_{21} = IR + L\frac{dI}{dt} + (\Phi_2 - \Phi_1) \qquad [2]$$

where

$$L = \int_1^2 \frac{\mu \bar{A} \cdot d\bar{l}}{I} \qquad [3]$$

Notice that in (3) the definition of L is somewhat different from that in Eq. 5.04(6). It should be noted that the definition (3) for inductance is good even for a non-continuous current path; but the concept of inductance based primarily on flux linkages would require special explaining to have meaning for such a case. To make the two definitions identical, despite the apparent difference, consider (3) as the general definition; then Eq. 5.04(6) is simply a modification of (3) in which there are no circuit discontinuities.

There remains a consideration of $(\Phi_2 - \Phi_1)$.

$$\Phi = \int_V \frac{[\rho]\, dV}{4\pi\epsilon r}$$

Retardation may be neglected as before for a circuit small compared with wavelength. Then, provided stray capacitance is negligible so that distributed charges on the surface of the wire are small compared with the charges concentrated at the discontinuity, the value of $\Phi_2 - \Phi_1$

is proportional to the total charge q at the discontinuity. Let $1/C$ be that constant of proportionality;

$$\Phi_2 - \Phi_1 = \frac{q}{C} \tag{4}$$

The charge at the discontinuity may also be related to current flowing toward the discontinuity by

$$q = \int I\, dt$$

So the circuit equation (2) finally becomes

$$V = IR + L\frac{dI}{dt} + \frac{1}{C}\int I\, dt \tag{5}$$

This equation is the usual equation written for a low-frequency series circuit containing resistance, inductance, and capacitance. The constant C of the factor of proportionality is thus identified directly with the low-frequency capacitance. In fact (4) constitutes the usual definition for capacitance of a condenser.

Problem 5.05. Demonstrate for a circuit with capacitive discontinuity that the definition of inductance based upon flux linkages may be applied to give a result essentially the same as that obtained from the general formula, Eq. 5.05(3), provided the discontinuity is small.

5.06 Summary of Approximations in Deriving the Approximate Circuit Equations

Starting from a rigorous equation, approximations have been made leading to the familiar circuit equations based upon Kirchhoff's laws, and the usual definitions of inductance and capacitance. These approximations were:

1. Current was assumed concentrated in a thin conducting filament of the same value everywhere along the path.

2. An effect from a change in current at one point in the circuit was assumed to be felt instantaneously at all other points of the circuit. Circuit dimensions, in other words, were assumed small compared with wavelength.

3. To arrive at the concept of capacitance, the discontinuity was assumed very small, and the definition of inductance was made more general so as not to require a completely continuous conducting path.

Of course, it should be recognized that there are practical, important cases where the circuit concept utilizing inductance and capacitance may be extended so that it is useful although not all the above approxi-

mations are justified. The extension of circuit theory to include distributed inductances and capacitances as well as lumped parameters is one example of this. Later, we shall consider also inductances in circuits in which the cross section of current flow is not of infinitesimal area, and in which the ideas of inductance and capacitance will still be of practical use.

HIGH-FREQUENCY OR LARGE-DIMENSION CIRCUIT CONCEPTS

5.07 Extension of the Circuit Inductance Concept

Of the approximations necessary to obtain the usual low-frequency definition of inductance, the assumptions of infinite velocities of propagation and negligible distributed capacities are most directly related to frequency. At the higher frequencies, the time required to propagate the effect of a change in charge or current over the dimensions of the circuit may be appreciable compared with a period of the changing effects. We shall next consider the circuit at such frequencies, assuming for the moment that distributed capacities are not yet of importance, so that the current still has effectively the same magnitude at all points about the circuit.

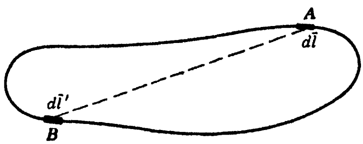


FIG. 5.07.

Consider the circuit of Fig. 5.07. With assumed infinite velocities of propagation, the effect of a change in current in the element $d\bar{l}$ at A would be felt instantaneously at all other points of the circuit (as at the element $d\bar{l}'$ at B), and for steady state sinusoidal time changes, only induced voltage drops in time quadrature with the current would be obtained at any point of the circuit. However, when finite velocities are considered, the time necessary to propagate the effect of a change in current at A to any other point B may be great enough so that this exact 90° relationship is destroyed. There may then result from the changing magnetic effects a component of induced field in phase with the current as well as an alteration in the magnitude of the 90° out-of-phase component. These corrections might be calculated relatively simply if the current were assumed to be all of the same phase and magnitude around the circuit; but it must be recognized that this need not always be true when retardation is of importance. Retardation enters because the time necessary to propagate an effect of changing current at one point of a circuit to another point through space is appreciable compared with the period of changing current; phase differences between currents at different points about the circuit enter because the time necessary

to propagate changing currents about the conductor of the circuit is appreciable compared with a period of changing current. The two effects are closely allied. And if current, at any one instant in time, varies around the loop, there must be a temporary piling up or a decreasing of charge at various points around the loop. So our suggestion to separate retardation effects and distributed charge effects appears of doubtful value.

However, the problem of phase differences and distributed charge is a bit difficult until we have had more experience with wave propagation along conductors, which is the subject of study in later chapters. We know already from the transmission line study of Chapter 1 (and a transmission line is an extreme example of a circuit large compared with wavelength) that there are often certain conditions under which the waves may combine to form a standing wave pattern with current everywhere in phase. Similarly, in certain other circuits large compared with wavelength, it will often be possible to have standing waves so that all points about these circuits have currents in phase. The following analysis, which considers retardation only, and assumes no phase differences, is directly applicable to such circuits. But we need not attach rigor to the following conclusions because the main objective is to indicate the importance of retardation alone in any circuit in a qualitative way whether the currents are in phase and of the same magnitude all around the circuit or not.

With the assumption of a thin filament of current of the same value and phase at all points about the circuit, Eq. 5.04(2) still holds, with vector potential given by Eq. 5.04(3).

$$V = IR + \frac{\partial}{\partial t}\oint \mu\bar{A}\cdot d\bar{l} \qquad [1]$$

$$\bar{A} = \oint \frac{[I]d\bar{l}'}{4\pi r} \qquad [2]$$

If the current is sinusoidal with time, it may be written

$$I = f(t) = I_0 e^{j\omega t}$$

Then the retarded value of current is

$$[I] = f\left(t - \frac{r}{v}\right) = I_0 e^{j\omega\left(t - \frac{r}{v}\right)}$$

Thus (1) becomes

$$V = IR + j\omega I_0 e^{j\omega t}\oint\oint \frac{\mu e^{-\frac{j\omega r}{v}}d\bar{l}'}{4\pi r}\cdot d\bar{l} \qquad [3]$$

From the equivalence,

$$e^{-jx} = \cos x - j \sin x$$

the integral in the above circuit equation may be broken into a real and an imaginary part.

$$V = IR + j\omega I \oint\oint \frac{\mu \cos\left(\frac{\omega r}{v}\right) \overline{dl}' \cdot \overline{dl}}{4\pi r} + I \oint\oint \frac{\mu\omega \sin\left(\frac{\omega r}{v}\right) \overline{dl}' \cdot \overline{dl}}{4\pi r} \qquad [4]$$

This equation might then be written as

$$V = I[(R + R_r) + j\omega L] \qquad [5]$$

where

$$L = \oint\oint \frac{\mu \cos\left(\frac{\omega r}{v}\right) \overline{dl}'}{4\pi r} \cdot \overline{dl} \quad \text{henrys} \qquad [6]$$

and

$$R_r = \oint\oint \frac{\mu\omega \sin\left(\frac{\omega r}{v}\right) \overline{dl}' \cdot \overline{dl}}{4\pi r} \quad \text{ohms} \qquad [7]$$

Analogy between (5) and the familiar low-frequency circuit equation using complex notation, identifies L as inductance, but from (6), L is seen now to be a function of frequency.[2] Its connection with low-frequency inductance, which is taken as a constant of geometry independent of frequency, becomes apparent if the cosine term in (6) is written in series form.

$$L = \mu \oint\oint \left(1 - \frac{\omega^2 r^2}{2!\, v^2} + \frac{\omega^4 r^4}{4!\, v^4} \cdots\right) \frac{\overline{dl}' \cdot \overline{dl}}{4\pi r} \qquad [8]$$

At low frequencies ($\omega r/v$ very small compared with unity) all terms but the first are negligible in the series, so that

$$L_{LF} = \oint\oint \frac{\mu \overline{dl}' \cdot \overline{dl}}{4\pi r} \qquad [9]$$

This is actually a well-known formula (Neumann's) for low-frequency inductance as a function of circuit geometry, and is, of course, independ-

[2] This is not to be confused with the change of inductance with frequency due to skin effect phenomena which is another matter and will be studied in the next chapter.

ent of frequency.[3] At higher frequencies, other terms of the series appear as correction terms to this low-frequency value of inductance.

5.08 Circuit Radiation Resistance

Let us investigate the additional term IR_r which appears in the circuit equation when the frequency is high. This term is in phase with the current, just as is the ohmic term IR, and so represents an actual departure of energy from the source. The ohmic term represents energy transfer from the source to heat in the conductors. The new in-phase term does not represent any such dissipation of electromagnetic energy into heat energy, but it does represent an actual energy which leaves the circuit and can accordingly be labeled radiated electromagnetic energy. In the chapter on radiation, several ways of looking at this radiation term and of calculating its magnitude will be studied. For the present, we are most interested in the term as a correction term to the circuit equations at high frequencies. From this point of view, it is convenient to expand Eq. 5.07(7) in its series form, as was done previously for Eq. 5.07(6).

$$R_r = \oint\oint \frac{\mu\omega}{4\pi r}\left(\frac{\omega r}{v} - \frac{\omega^3 r^3}{3!\, v^3} + \frac{\omega^5 r^5}{5!\, v^5}\cdots\right) \bar{dl}' \cdot \bar{dl} \qquad [1]$$

But since $\oint \bar{dl}' = 0$, the first term (which contains no r) disappears entirely and there remains only

$$R_r = \oint\oint \frac{\mu}{4\pi}\left(-\frac{\omega^4 r^2}{3!\, v^3} + \frac{\omega^6 r^4}{5!\, v^5}\cdots\right) \bar{dl}' \cdot \bar{dl} \qquad [2]$$

R_r, because of its similarity to the ohmic term, may be called a *radiation resistance.*

This radiation resistance, representing an in-phase component of the induced voltage due to varying magnetic effects, and the correction to inductance or out-of-phase component of induced voltage found in the previous article, were direct consequences of retardation. We know, of course, that there are other factors which have not been considered, so that expressions given for these correction terms should not be taken as rigorous. Frequently the changes in current about the path due to stray capacity and the differences in phase about the path, both of which

[3] Actually Neumann's formula is of practical use only in calculating mutual inductance, since it has in it the assumption of filamentary currents, and this leads to bothersome infinities in evaluation of the self inductance. All this will be discussed in detail in the next chapter when the objective will be the computation of circuit impedance.

have been neglected, may be more important than the retardation which was considered. In spite of this incompleteness, the foregoing analysis does demonstrate the important effects of retardation.

5.09 Example of Use of Circuit Concepts in an Inductive Circuit of Large Dimensions

When retardation is included in the analysis of an inductive circuit, it has been found that a radiation of energy term may be expected. We say " may " rather than " must " because so far it has been shown only that the electric field component induced by a current element of the circuit will not always be 90° out of phase with that current element. If the net electric field due to all the circuit's current happens to have a component in phase with the current at the point of the circuit where the electric field is being evaluated, then there will indeed be radiation. It is nevertheless possible to imagine situations (one of which will be met later in the analysis of the ideal transmission line) in which, although retardation is included, the electric field still is 90° out of phase with the current at that point. This happens because the current itself is not uniform in phase and magnitude around the circuit. There are many other cases where this beautiful compensation does not take place. One such is a circle of current where symmetry makes it possible to have constant current around the loop.

Consider a circular loop of wire having a radius a (Fig. 5.09). The magnitude of $d\bar{l}'$ is $a d\phi$. From the definition of the scalar product,

$$d\bar{l}' \cdot d\bar{l} = a\, d\phi\, dl \cos\phi$$

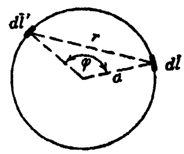


FIG. 5.09. Circular current-carrying loop.

The distance between the elements $d\bar{l}$ and $d\bar{l}'$ is

$$r = 2a \sin\frac{\phi}{2}$$

Thus, neglecting all but the first term in the expression for radiation resistance, Eq. 5.08(2),

$$R_r = -\frac{\mu\omega^4}{24\pi v^3}\oint\oint r^2\, d\bar{l}' \cdot d\bar{l}$$

$$= -\frac{\mu a^3\omega^4}{6\pi v^3}\int_0^{2\pi a} dl \int_0^{2\pi} \sin^2\frac{\phi}{2}\cos\phi\, d\phi$$

$$R_r = \frac{\mu\pi a^4\omega^4}{6v^3}$$

$$= \frac{\eta\pi}{6}\left(\frac{2\pi a}{\lambda}\right)^4 \quad \text{ohms}$$

where

$$\lambda = \frac{v}{f} = \frac{2\pi v}{\omega} = \frac{2\pi}{\omega\sqrt{\mu\epsilon}}$$

$$\eta = \sqrt{\frac{\mu}{\epsilon}}$$

For free space surrounding the wire, $\eta_0 = 120\pi$ ohms, and

$$R_r = 20\pi^2 \left(\frac{2\pi a}{\lambda}\right)^4 \quad \text{ohms}$$

As a numerical example, consider a loop with circumference one quarter of a wavelength,

$$R_r = 20\pi^2 (\tfrac{1}{4})^4 = 0.773 \quad \text{ohms}$$

Problem 5.09(*a*). Obtain in terms of ratio of loop circumference to wavelength the first correction term to inductance by Eq. 5.07(8) for the above circular loop of wire.

Problem 5.09(*b*). The inductance of a certain single loop of wire is five microhenrys. Determine the ratio of wire circumference to wavelength for which the above correction term is 0.1 microhenry.

Problem 5.09(*c*). Suggest a way to "drive" the circular loop of Fig. 5.09 so as to produce an applied electric field, equal in phase and magnitude everywhere around the loop, a condition that would yield the assumed uniform current distribution.

5.10 Extension of the Circuit Capacitance Concept

We have seen that the term in the exact circuit equation 5.03(1), which gives voltage drop in terms of changing magnetic effects, reduces to the ordinary inductance drop at low frequencies, but is really more general. At high frequencies, finite velocities of propagation may be important enough so that the value of the reactive drop from this term is changed, and a new radiated term in phase with the current flow is appreciable. It remains to examine similarly the term which gives voltage drop in terms of charges. It is always true that the line integral of the gradient of Φ must be zero around any closed path, so that a circuit with discontinuity must be selected and the integration carried only to the terminals of the discontinuity (capacitors) if the term is to appear at all. Consideration of such a discontinuity led directly to the usual concept of capacitance for low-frequency circuit effects in Art. 5.05.

Consider the circuit of Fig. 5.10. Charges are again considered as concentrated at the discontinuity. That is, we are still neglecting any stray capacity effects from distributed charges resting on the surface

of the wire. At low frequencies, the electric field at any point, such as point P, from these two lumped charges, is always 90° out of time phase with the current, so that this component represents no average power flow. The energy in the electric field simply oscillates between the source and the surrounding space. If the frequency becomes very high, the length of time necessary to propagate the effect of charges $+q$ and $-q$, at 2 and 1, to P, any other point on the circuit, may destroy the previous 90° phase relationship. It is quite apparent that there will now be a component of the electric field due to charges which is in phase with the current at any point of the circuit because of this retardation, quite similar to that found for changing magnetic effects, and also representing a radiation of energy.

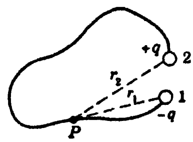


FIG. 5.10.

5.11 Considerations Involved in an Exact Approach to Circuits of Large Dimensions

At the higher frequencies, stray capacities become of increasing importance, and the assumptions of uniform current distributions and no charges on the surfaces of wires may require revision if useful answers are to be predicted. The exact circuit equation is always that derived in Art. 5.03.

$$V_{21} = \int_1^2 \frac{\bar{\imath}}{\sigma} \cdot \overline{dl} + \int_1^2 \mu \frac{\partial \bar{A}}{\partial t} \cdot \overline{dl} + \int_1^2 \nabla\Phi \cdot \overline{dl} \qquad [1]$$

with

$$\bar{A} = \int_V \frac{[\bar{\imath}]\, dV}{4\pi r} \qquad [2]$$

$$\Phi = \int_V \frac{[\rho]\, dV}{4\pi\epsilon r} \qquad [3]$$

Difficulties in applying these equations arise since the current and charge distributions are not known, but are determined by the field distributions which are calculated from the retarded potentials which depend upon current and charge distribution — a vicious circle! The exact solution of this problem is usually of prohibitive difficulty. However, it is often possible to assume a reasonable current distribution, calculating from it the retarded potentials and hence the fields; from these the first assumption of current distribution may be corrected, and the process repeated until the desired accuracy is reached. This is not often done, however, since it is obviously a laborious method, and as

soon as frequencies become so high that these distributed capacities are important, transmission line theory usually offers a superior way of looking at the problem. So we shall talk more about what might be called circuits of large dimensions later under the various headings of transmission lines, wave guides, resonant cavities, and antennas.

In calculating the radiation of energy by the use of circuit equations, it may be necessary only to consider one step of the above outlined procedure; that is, reasonable current distributions are assumed and radiated energy is calculated from these. If the concept of radiation resistance is used, it must be defined properly, and used with care, for it is the total energy radiated from the system which has meaning. If it is desired to express this radiation by multiplying some radiation resistance by the square of a current, it must be remembered that current may no longer be the same at all points around the circuit. Thus the value of radiation resistance for a given system will depend upon the particular current which is selected for this purpose. Actually a radiation resistivity could be defined for every point of the system in such a way that total energy radiated from the system might be obtained by an integration of this radiation resistivity multiplied by the square of current density over all the circuit. It is total radiation resistance rather than radiation resistivity which is most often used by radio engineers. It will be considered more thoroughly in the later chapter on radiating systems.

5.12 Self-Enclosing a Circuit to Prevent Radiation

When retardation is neglected in the analysis of a circuit, the result will inevitably contain no possibility for radiation of energy. When retardation is included, then the possibility exists that the answer may disclose loss of energy by electromagnetic field leakage (or radiation) into the surrounding space. This does not mean that retardation of itself, no matter how great, always leads to radiation.

The emphasis should rather be on the fact that retardation means that the electric fields arising from any current element will, in the surrounding space, not be 90° out of phase with that current element. There is accordingly a possibility always that the total induced electric field at any point in the circuit may have a component in phase with the current at that point. This possibility passes to actuality in circuits of the type discussed so far, in which the current flows in filamentary paths. The validity or limits of the common circuit concepts were established on the basis of such circuits; but now that this is done, there is no reason why we must limit ourselves to these circuits. Useful and well-nigh universal though they may be at low frequencies, it is apparent that this

loss characteristic, radiation, that rises rapidly with frequency, will limit their application. For antennas, devices that are chosen because they do indeed "leak" electromagnetic energy to the surrounding space, these circuits are candidates. For narrow-band filters, resonant circuit impedances, and a host of other conventional circuit applications, it is much preferred to have a non-radiating circuit.

Now there are ways to minimize radiation, even to make it zero for all practical purposes. We mention two of them here, not to complete the discussion of circuits (because to appreciate these new things fully will require techniques of electromagnetic waves that will be discussed in following chapters) but rather to make it evident that the story on circuits is not complete in this chapter, although its purpose was to examine circuit concepts. One way to prevent radiation is to enclose completely the circuit and source by a very good conductor. Such a shield, as we shall see in detail later, will stop the electromagnetic energy leaving the circuit, reflect it, and cause additional electric field in the circuit that will buck out that undesired induced electric field component in phase with the current at each point of the circuit. Practically, the conductivity of the shield cannot be infinite, and so some small amount of energy will get through, and the reflected electric field will fall short of exactly neutralizing the in-phase component of induced electric field. We cannot discuss this problem completely until we have learned more about handling the electromagnetic energy as a wave phenomenon.

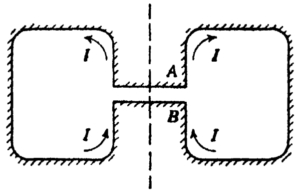


FIG. 5.12. A self-enclosed circuit.

Another way to build a circuit so as to minimize radiation is exemplified in Fig. 5.12. Here we deal with cavities, such that for any cross section which includes the axis of symmetry we have a circuit of the parallel resonant L–C type, consisting of the condenser plates A and B closed by the one turn inductance which encloses itself and the condenser. Because of the fact (which emerges easily from electromagnetic wave studies) that all electromagnetic effects, practically speaking, fail to penetrate metals at very high frequencies, the leakage by radiation from such a circuit will be negligible. This means that current distribution is not uniform but rather such that the net induced electric field at every point in the conductor is essentially 90° out of phase with the current at that point even at very high frequencies when retardation must be and is included. Such a circuit is best analyzed as a resonant cavity by the use of electromagnetic wave pictures and equations, so we shall leave the discussion at this point. We add only that certain concepts which can

properly be called " circuit concepts " will be helpful in studying these resonant cavities and the fundamental concepts discussed in this chapter should not be forgotten when we pass over to the wave method.

5.13 Mutual Couplings

In the circuit theory so far only a single mesh was considered. It was shown how the electromagnetic equations reduce for the low-frequency case to the familiar circuit equations developed from the single Kirchhoff's law for such a single-mesh circuit. For higher frequencies it was shown how the retardation effects introduce corrections to the values of inductance and capacitance to be used in the circuit equations,

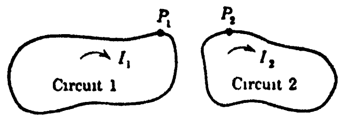


FIG. 5.13a. Two circuits coupled through inductive effects.

and may add in-phase terms representing radiation of energy from the loop. Until now the very important phase of classical circuit theory involving couplings between different circuit meshes has been avoided. This coupling may be from mutual inductances, from capacities whose fields influence one another, or from conduction links.

If the coupling is of the form of a mutual inductance, Fig. 5.13(*a*), there is at every point a contribution to the vector potential from the currents in both of the loops. For each mesh, the integral circuit equation is of the form of Eq. 5.03(1),

$$V = \oint \frac{\bar{\imath}}{\sigma} \cdot d\bar{l} + \frac{\partial}{\partial t} \oint \mu \bar{A} \cdot d\bar{l} + \oint \nabla \Phi \cdot d\bar{l} \qquad [1]$$

but in calculating the value of $\bar{A}$ to use for any point P_1 in circuit 1, the current I_2 must be considered with proper retardation just as in the case of the current I_1. Induced field at P_1 from this changing magnetic effect will in general have a component in phase with current due to retardation, as well as the usual 90° out-of-phase mutual inductance component, thus causing additional radiation of energy. Similarly, in calculating the value of $\bar{A}$ to use at any point P_2 of circuit 2, the current I_1 with proper retardation must be used as well as I_2.

If the coupling is from charges (usually thought of as a capacitive coupling) rather than, or in addition to, the inductive coupling from currents, a similar effect arises. Thus to calculate the value of Φ to use in the circuit equation for any point P_1 in circuit 1 of Fig. 5.13*b*, contri-

butions from $+q_2$ and $-q_2$ of circuit 2 must be considered as well as $+q_1$ and $-q_1$ of circuit 1, with proper retardations for each. In general, the induced field from these charges of circuit 2 will contribute an additional component of field in phase with current because of retardation, thus presenting further possibility of energy radiation. The situation illustrated by Fig. 5.13*b* is that of condensers in the two meshes insufficiently shielded to prevent their fields from influencing one another.

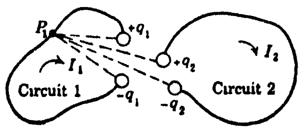


FIG. 5.13*b*. Two circuits coupled through capacitive effects.

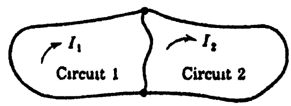


FIG. 5.13*c*. Two circuits coupled conductively.

If the coupling between meshes is of a conductive nature, Fig. 5.13*c*, the general circuit equation holds about each path, and separate equations, each of the form of (1), would be written for each mesh. At any point the value of $\bar{A}$ and Φ used must be calculated for all charges and currents of both meshes, and continuity of charge must be applied at the junction, the second Kirchhoff law.

For any of the above methods of coupling between meshes, the electromagnetic field approach would reduce at low frequencies to the usual classical circuit equations derived from Kirchhoff's laws. At high frequencies, the approach is as outlined here, but it must be recognized that it is very difficult to apply except for the simplest of configurations. The value of the low-frequency circuit approach is emphasized all the more by these difficulties, and it is consequently important to recognize its assumptions and limitations. Such a true understanding is necessary in extending the valuable circuit approach to cases where not all these assumptions are strictly justified.

6

SKIN EFFECT AND CIRCUIT IMPEDANCE ELEMENTS

6.01 Introduction

The discussion of the previous chapter justified circuit concepts, like inductive and capacitive reactances, as they are universally used and understood in conventional circuit analysis. To be sure, it was found that the well-known definitions of all these quantities involve approximations; but, on the whole, for circuits physically small compared with wavelength, rigorous approach by Maxwell's equations shows that the ordinary methods of circuit analysis stem from correct formulations and are accurate. Even at higher frequencies, where the approximations become poorer, these circuit concepts are still practical for part if not all the problem, although correction terms may have to be used.

Chapter 5 also discussed the various factors which enter into a rigorous analysis of all the effects which may take place in the neighborhood of a simple loop of conductor. We shall be satisfied in the present chapter to make simplifying assumptions, particularly those approximations which permit us to think in terms of an applied voltage around a circuit being taken up in impedance drops: ohmic resistance, inductive reactance, and capacitive reactance. There is a tremendous range of practical problems for which these assumptions are justified, and when the approximate quantities called resistance, capacitance, and inductance have been computed, the electromagnetics of these circuits may be said to have been completely worked out. The obtaining of these impedance elements will occupy us throughout this chapter.

Of course, a huge store of knowledge exists on the handling of circuit problems once all the equations are set up and the circuit parameters computed. This special subject of circuit analysis and synthesis is not, however, within the scope of interest of this text and the reader is left to consult the numerous sources dealing primarily with such material.

SKIN EFFECT AND THE INTERNAL IMPEDANCE OF A CONDUCTOR

6.02 The Importance of Skin Effect in Impedance Calculations

Many aspects of a phenomenon called *skin effect* will be important throughout the book. This chapter will be concerned mainly with appli-

cations to impedance calculations for wires and other conductors, yet it is important to start with a much broader picture of the subject. Skin effect is most often introduced through the example of high-frequency current flow in a solid round conductor, in which it can be demonstrated that current flow at very high frequencies is essentially concentrated in a thin layer or skin near the surface. Students often leave such first introductions to the subject with the impression that this is the most important aspect of skin effect, and worse, believe erroneously that the above phenomenon is caused by some sort of a mutual repulsion between small filamentary current elements in the wire.

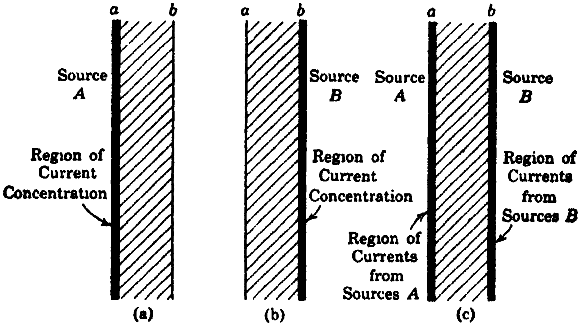


FIG. 6.02*abc*. Current concentration caused by skin effect.

With such a picture, they expect always to find current seeking the outside of any conducting system, an impression that would be unfortunate indeed as a preliminary to the study of resonant cavities, wave guides, shields, etc., in which currents may be concentrated on the inner, not outer, walls of the conductors.

A broad picture of skin effect shows that it is a phenomenon which tends to concentrate currents on the surfaces of conductors that are nearest to the field sources producing them. Thus in Fig. 6.02*a*, if there is an exciting source A of extremely high frequency near a conducting sheet, current may be essentially concentrated on the side a of the sheet; if there is a source B, as in Fig. 6.02*b*, current may be concentrated on the side b; if there are both sources, A and B, Fig. 6.02*c*, there may be currents on the two sides, a due to A and b due to B, for all practical purposes completely independent. At such frequencies the conducting wall has acted as a complete shield between the sources A and B. The reservations "essentially" and "for all practical purposes" will be clearer when we next study the equations of skin effect. It will then be seen that penetration of current into the conductor decreases gradually, so that current is not actually concentrated in a small layer at the surface with no current beneath. In any real conduc-

tor, current will not actually decrease to zero no matter how thick the conductor. This statement should not mask the fact that at the highest radio frequencies current density may decrease to one millionth its surface value in a distance of only a few thousandths of an inch, so that any practical thickness of any conductor becomes the " practically perfect " shield referred to above.

The general reason for skin effect behavior can be visualized in terms of applied and induced voltages. Thus imagine a high-frequency source as in Fig. 6.02*d*, producing an applied electric field E_0 in the neighborhood of the conductor. This must cause current flow in the conductor, producing a magnetic field at right angles to E_0. This changing magnetic field produces an induced electric field E' opposite to E_0. If a study is made of the two closed line integrals, 1–2–3–4–1 and 1–2′–3′–4–1, it is found that more magnetic flux is enclosed in the latter, so that induced voltage around this path is the greater, and it may be deduced that the induced field along 2′–3′ is greater than that along 2–3. It follows that there is less net field, $E_0 + E'$, left to produce current flow as one progresses farther into the conductor.

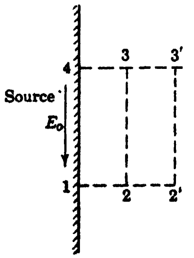


Fig. 6.02*d*.

Another physical picture of skin effect phenomena will follow from the wave concepts of Chapter 7. From such a viewpoint, one can consider the source as a source of waves which impinge upon the conductor. Some wave energy is, of course, reflected due to impedance mismatch at the discontinuities between air and conductor. Those waves which pass into the conductor attenuate at a rate determined by the conductivity of the conductor, just as transmission line waves attenuate in a line with high leakage conductance.

The classical example of current flow in a solid round conductor is now seen to be a special case of this general viewpoint; certainly, if the conductor is solid, the exciting sources must be on the outside, so current will concentrate near the outside. However, if exciting sources are on the inside of a hollow conductor, as they are for the outer conductor of a coaxial line, current will concentrate on the inner wall of that conductor. Finally, we might imagine a double coaxial line, as in Fig. 6.02*e*, formed of good conductors and operated at very high frequencies. Currents due to the source A are concentrated on the walls a and a'; currents due to B are concentrated on the walls b and b'. For all practical purposes shielding between the two coaxial regions may be considered as perfect, and phenomena of the two regions are completely independent.

The general concentration of current into thin layers, as found in skin effect phenomena, should have a marked effect on impedances, causing them to change with frequency. If current is concentrated over a smaller part of the cross section of a conductor than at low frequencies, the effective conductor cross section is decreased and resistance should increase. Also, if penetration of fields into the conductor becomes less as frequency increases, there should not be as much magnetic flux inside the conductor and internal inductance should decrease. All these phenomena will be studied quantitatively in articles to follow.

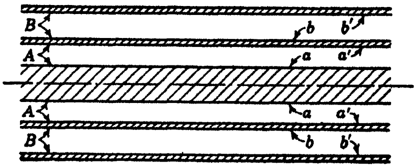


Fig. 6.02e.

6.03 Equation Determining Current Distribution in a Conductor

For a quantitative discussion of the impedance of conductors as a function of the current distribution there will first of all be required an equation which comes from Maxwell's equations and which contains only the current density and the coordinates.

Maxwell's equations are:

$$\nabla \cdot \bar{D} = \rho \qquad [1]$$

$$\nabla \cdot \bar{B} = 0 \qquad [2]$$

$$\nabla \times \bar{H} = \bar{\imath} + \frac{\partial \bar{D}}{\partial t} \qquad [3]$$

$$\nabla \times \bar{E} = -\frac{\partial \bar{B}}{\partial t} \qquad [4]$$

In addition there is Ohm's law, which may be taken as the definition of a conductor:

$$\bar{\imath} = \sigma \bar{E}$$

where the constant σ is the conductivity of the conductor. The current density appearing in Maxwell's equations may now be expressed in terms of the fields. Not much has yet been said about the charge density ρ in conductors, but an elimination between (3) and (1) makes it possible to obtain an equation in ρ alone which should provide information about it. Equation (3) may be rewritten with the aid of Ohm's law to eliminate the current density vector.

$$\nabla \times \bar{H} = \sigma \bar{E} + \frac{\partial \bar{D}}{\partial t} \qquad [5]$$

We now take the divergence of both sides of (5) and recall that the divergence of the curl of any vector is identically zero.

$$\nabla \cdot \nabla \times \bar{H} \equiv 0 = \sigma\left(\frac{\nabla \cdot \bar{D}}{\epsilon}\right) + \frac{\partial(\nabla \cdot \bar{D})}{\partial t}$$

Divergence of $\bar{D}$ is given in terms of charge density by the first of Maxwell's equations. Substituting,

$$\frac{\sigma}{\epsilon}\rho + \frac{\partial \rho}{\partial t} = 0 \qquad [6]$$

Equation (6) is the desired differential equation in ρ alone. Its solution,

$$\rho = \rho_0 e^{-\frac{\sigma}{\epsilon}t} \qquad [7]$$

shows that any charge density which might exist in a conductor obeying Ohm's law must decay exponentially, and at an extremely rapid rate in good conductors. Any charges, if ever placed in the interior of such a conductor, would flow at once to the surface; consequently, unless we quickly think up some means of continuously generating free charges in the interior of the metal, the free charge term in Maxwell's equation must be zero for all steady state conditions. So, in good conductors,

$$\nabla \cdot \bar{D} = 0 \qquad [8]$$

This result was a direct consequence of applying Ohm's law, requiring that current be directly proportional to electric field. Not all currents are of this type. For example, there is the space charge type of current that consists entirely of a motion of free charges in space. But currents in conductors, in so far as Maxwell's equations and Ohm's law are correctly used to describe them, will not be accompanied by the presence of charge density in the conductors.

It should be clear to the student that in stating effortlessly that the charge density is zero in a conductor, and that the current flows anyway, thereby implying that conduction is due to a transfer of charge from atom to atom, rather than a flow of free charges, we are not covering the whole of the theory of conduction. In fact, this question is pretty well avoided. If Ohm's law and Maxwell's equations are true, then the conclusions we draw from them are true. This does not rule out the presence of various charges in a conductor, but it does say that they must in effect add up to zero at every point. As to what goes on inside a conductor that makes $\bar{\imath}$ proportional to $\bar{E}$, that is being left to the experts in modern theory of solids.

Equation (8) may be derived more easily perhaps by limiting the dis-

cussion to steady state sinusoidal variations with time. If all quantities are proportional to $e^{j\omega t}$ then (3) becomes

$$\nabla \times \bar{H} = (\sigma + j\omega\epsilon)\bar{E} \tag{9}$$

Again taking the divergence of both sides

$$\nabla \cdot \nabla \times \bar{H} = 0$$

and thus

$$\nabla \cdot (\sigma + j\omega\epsilon)\bar{E} = (\sigma + j\omega\epsilon)\, \nabla \cdot \bar{E} = 0$$

or

$$\nabla \cdot \bar{E} = \nabla \cdot \bar{D} = 0$$

The equations may be further simplified, since the displacement current will never be appreciable in any reasonably good conductor, even at the highest radio frequencies. Consider again variations which are sinusoidal with time (of the form $e^{j\omega t}$). The terms to be compared in (9) are σ and $\omega\epsilon$. The precise values of ϵ for conductors are not known, yet most indications show that the range of dielectric constants is much the same for conductors as for dielectrics. For platinum, a relatively poor conductor, the term $\omega\epsilon$ becomes equal to σ at about 1.5×10^{15} cps, if the dielectric constant is taken as ten times that of free space. This frequency is in the range of ultraviolet light. Consequently, for all but the poorest conductors (such as earth) the displacement current term is completely negligible compared with conduction current at any radio frequency.

There remains

$$\nabla \times \bar{H} = \sigma\bar{E} \tag{10}$$

The curl of both sides may be taken, and the left side expanded.

$$\nabla \times \nabla \times \bar{H} \equiv \nabla(\nabla \cdot \bar{H}) - \nabla^2\bar{H} = \sigma \nabla \times \bar{E}$$

Values for $\nabla \cdot \bar{H}$ and $\nabla \times \bar{E}$ are obtained from Maxwell's equations (2) and (4), leaving

$$\nabla^2\bar{H} = \sigma\mu \frac{\partial \bar{H}}{\partial t} \tag{11}$$

This equation for the variation of $\bar{H}$ in a conductor is in the form of a standard differential equation similar to Laplace's equation, or the wave equation. The equation is often called the skin effect or distribution equation and may also be derived in terms of $\bar{E}$, taking first the curl of (4) instead of (3), and expanding as before to yield

$$\nabla^2\bar{E} = \sigma\mu \frac{\partial \bar{E}}{\partial t} \tag{12}$$

Since $\bar{\imath} = \sigma\bar{E}$, the same equation may be written in terms of current density.

$$\nabla^2\bar{\imath} = \sigma\mu \frac{\partial \bar{\imath}}{\partial t} \quad [13]$$

When all quantities can be regarded as varying as $e^{j\omega t}$, the above equations may be written

$$\nabla^2\bar{H} = j\omega\sigma\mu\bar{H} \quad [14]$$

$$\nabla^2\bar{E} = j\omega\sigma\mu\bar{E} \quad [15]$$

$$\nabla^2\bar{\imath} = j\omega\sigma\mu\bar{\imath} \quad [16]$$

These equations give the relation between space and time derivatives of magnetic field, electric field, or current density at any point in a conductor. It remains to solve these differential equations subject to the boundary conditions imposed by certain physical shapes of interest for practical conductors.

6.04 Current Distribution in a Wire of Circular Cross Section

Most common of the conductors used in electrical circuits are round wires, wires of circular cross section. If the round wire forms a conducting path with no very sharp curvatures, as in many circuit applications, any small portion, or at least a differential length, may be treated as a straight circular cylinder. The current distribution may be computed exactly in such a straight cylindrical wire. This is only a first step in the problem, for the final interest in the current distribution lies in its effect upon impedance of the wire, that is, the resistance and inductance of the circuit containing the wire.

Before applying the distribution equation, let us be reminded that a distribution should be expected at high frequencies in which the greatest percentage of the current is concentrated near the outside of the wire. This follows from the reasoning of Art. 6.02. Since the sources must of necessity be applied outside of the solid wire, the current flow in response to an applied field induces an increasingly large back field as one progresses into the conductor, leaving less total electric field to produce current flow. As frequency increases, this effect becomes more pronounced, since rate of change of flux, and hence induced voltage, increases.

It will be assumed that external conditions are applied so that there need be no circumferential variations about the wire. If the wire is short compared with wavelength, there are only axial currents and no variations in the axial direction. The current distribution equation for

the axial current, written in cylindrical coordinates with no ϕ or z variations is then (Art. 3.12)

$$\frac{d^2 i_z}{dr^2} + \frac{1}{r}\frac{di_z}{dr} = j\omega\mu\sigma i_z$$

$$\frac{d^2 i_z}{dr^2} + \frac{1}{r}\frac{di_z}{dr} + T^2 i_z = 0 \qquad [1]$$

where

$$T^2 = -j\omega\mu\sigma$$

or

$$T = j^{-\frac{1}{2}}\sqrt{\omega\mu\sigma} \qquad [2]$$

A direct comparison of (1) with Eq. 3.15(1) shows that both have exactly the form of the zero order Bessel equation, although T is complex. A complete solution may be written, as in Eq. 3.16(8).

$$i_z = AJ_0(Tr) + BH_0^{(1)}(Tr) \qquad [3]$$

For a solid wire, $r = 0$ is included in the solution, and then it is necessary that $B = 0$ since a study of $H_0^{(1)}(Tr)$ shows that this would become infinite at $r = 0$, even though T is complex. Therefore,

$$i_z = AJ_0(Tr) \qquad [4]$$

The arbitrary constant A may be evaluated in terms of current density at the surface. Let

$$i_z = i_0 \quad \text{at} \quad r = r_0$$

Then from (4)

$$A = \frac{i_0}{J_0(Tr_0)}$$

and

$$i_z = \frac{i_0 J_0(Tr)}{J_0(Tr_0)} \qquad [5]$$

T is complex and it may seem troublesome to find a Bessel function of a complex quantity. However, as in cases where we are confronted with sines, cosines, and exponentials of complex quantities, we can resort to the power series definition for the proper function. Referring to the power series for J_0 it is seen that the function will have both real and imaginary parts if the argument is complex. These may be calculated separately. Define

$$Ber\ (v) \equiv \text{Real part of } J_0(j^{-\frac{1}{2}}v)$$

$$Bei\ (v) \equiv \text{Imaginary part of } J_0(j^{-\frac{1}{2}}v)$$

That is,

$$J_0(j^{-1/2}v) \equiv Ber\ (v) + jBei\ (v) \qquad [6]$$

Ber (v) and *Bei* (v) are tabulated in many references.[1] Using these definitions and (2), (5) may be written:

$$\dot{i}_z = \dot{i}_0 \frac{Ber\ (r\sqrt{\omega\mu\sigma}) + jBei\ (r\sqrt{\omega\mu\sigma})}{Ber\ (r_0\sqrt{\omega\mu\sigma}) + jBei\ (r_0\sqrt{\omega\mu\sigma})} \qquad [7]$$

$$\left|\frac{\dot{i}_z}{\dot{i}_0}\right| = \left[\frac{Ber^2\ (r\sqrt{\omega\mu\sigma}) + Bei^2\ (r\sqrt{\omega\mu\sigma})}{Ber^2\ (r_0\sqrt{\omega\mu\sigma}) + Bei^2\ (r_0\sqrt{\omega\mu\sigma})}\right]^{1/2} \qquad [8]$$

$$\text{Phase of } \frac{\dot{i}_z}{\dot{i}_0} = \tan^{-1}\frac{Bei\ (r\sqrt{\omega\mu\sigma})}{Ber\ (r\sqrt{\omega\mu\sigma})} - \tan^{-1}\frac{Bei\ (r_0\sqrt{\omega\mu\sigma})}{Ber\ (r_0\sqrt{\omega\mu\sigma})} \qquad [9]$$

Let us rewrite the above equations in terms of a new parameter, called *depth of penetration* for reasons to be discussed in the next article and defined by

$$\delta = \frac{1}{\sqrt{\pi f\mu\sigma}} \qquad [10]$$

Then (6) and (7) are

$$\left|\frac{\dot{i}_z}{\dot{i}_0}\right| = \left[\frac{Ber^2\left(\frac{\sqrt{2}\,r}{\delta}\right) + Bei^2\left(\frac{\sqrt{2}\,r}{\delta}\right)}{Ber^2\left(\frac{\sqrt{2}\,r_0}{\delta}\right) + Bei^2\left(\frac{\sqrt{2}\,r_0}{\delta}\right)}\right]^{1/2} \qquad [11]$$

$$\text{Phase of } \frac{\dot{i}_z}{\dot{i}_0} =$$

$$\tan^{-1}\left[\frac{Ber\left(\frac{\sqrt{2}\,r_0}{\delta}\right)Bei\left(\frac{\sqrt{2}\,r}{\delta}\right) - Ber\left(\frac{\sqrt{2}\,r}{\delta}\right)Bei\left(\frac{\sqrt{2}\,r_0}{\delta}\right)}{Ber\left(\frac{\sqrt{2}\,r_0}{\delta}\right)Ber\left(\frac{\sqrt{2}\,r}{\delta}\right) + Bei\left(\frac{\sqrt{2}\,r_0}{\delta}\right)Bei\left(\frac{\sqrt{2}\,r}{\delta}\right)}\right] \qquad [12]$$

Plots of current densities as functions of radius in a round wire are shown in Fig. 6.04. Actually the magnitude of the ratio of current density to that at the outside of the wire is plotted as a function of the ratio of radius to outer radius of wire, for different values of the parameter (r_0/δ). Also for purposes of the physical picture, these are

[1] Dwight, "Tables of Integrals," Macmillan, 1934. Mclachlan, "Bessel Functions for Engineers," Oxford, 1934.

interpreted in terms of current distribution for a 1-mm copper wire at different frequencies by the figure in parentheses.

6.05 Current Distribution in a Flat or Plane Conductor; Depth of Penetration

The current distribution plots of Fig. 6.04 show, as predicted, that at the highest frequencies, most of the current is concentrated in a thin region near the wire surface. For these cases the curvature of the wire seems unimportant, and it might be expected that the distribution could be predicted accurately enough by taking out a small portion of the surface, neglecting the curvature, and analyzing it as a plane conductor of infinite depth. This is a convenient analysis since it may be applied to conductors of many other shapes, so long as curvatures of the surface are large compared with the depth over which most of the current is concentrated. We shall consequently next seek the current distribution in a plane conductor of infinite depth, with no current variations with width or length. This plane conductor of infinite length, infinite width, and infinite depth below the surface is described mathematically as a semi-infinite solid.

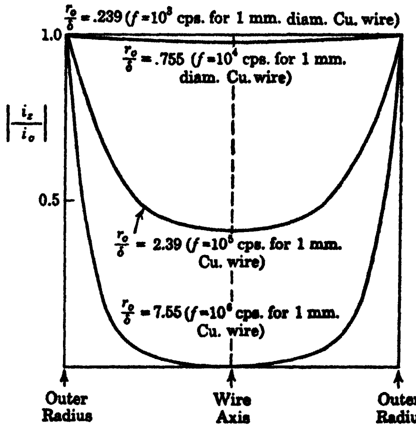


FIG. 6.04. Current distribution in cylindrical wire for different frequencies.

For the semi-infinite conductor, take the direction of current flow as the z direction, the normal to the surface as the x direction. It is assumed that there are no variations in the y or z direction. The current distribution equation is then only

$$\frac{d^2 i_z}{dx^2} = j\omega\mu\sigma i_z = \tau^2 i_z$$

where

$$\tau^2 = j\omega\mu\sigma \quad \text{or} \quad \tau = (1 + j)\sqrt{\pi f \mu\sigma} \qquad [1]$$

A complete solution to this equation is

$$i_z = C_1 e^{-\tau x} + C_2 e^{\tau x} \qquad [2]$$

Current density would increase to the impossible value of infinity at $x = \infty$ unless C_2 is zero. C_1 may be written as the current density at

the surface if we let $\dot{i}_z = \dot{i}_0$ when $x = 0$. Then

$$\dot{i}_z = \dot{i}_0 e^{-\tau x} \tag{3}$$

If use is made of the quantity δ defined by Eq. 6.04(10), the above may be rewritten

$$\frac{\dot{i}_z}{\dot{i}_0} = e^{-(1+j)\frac{x}{\delta}} = e^{-\frac{x}{\delta}} e^{-j\frac{x}{\delta}} \tag{4}$$

In this form it is apparent that magnitude of current decreases exponentially with penetration into the conductor, and δ has significance as the depth at which current density has decreased to $1/e$ (about 36.9 per cent) of its value at the surface. The phase of current also changes with increasing depth into the conductor according to the factor $e^{-j\frac{x}{\delta}}$.

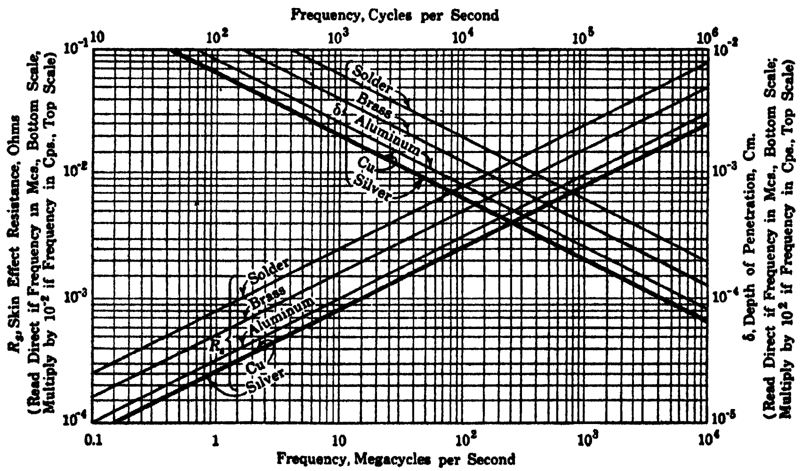


FIG. 6.05*a*. Skin effect quantities for plane conductors.

It is important to develop a familiarity for values of depth of penetration for different materials at different frequencies, and a study of the chart, Fig. 6.05*a*, is helpful for this. The student should also retain these facts:

1. Depth of penetration is smaller the better the conductor, the higher the permeability, and the higher the frequency, since it is inversely proportional to the square root of each of these.

2. Current does not fail to penetrate below the depth δ; this is merely the point at which current densities and fields have decreased to $1/e$ their value at the surface. Later another important significance of this quantity will appear.

3. The concept as stated here applies strictly only to plane solids. However, it may be extended to conductors of other shapes so long as the value of δ calculated is much smaller than any curvatures of the surfaces.

As an example of the last point, it is possible to study high-frequency current distribution in the solid round conductor by extending the plane analysis. The coordinate x, distance below the surface, is $(r_0 - r)$ for a round wire. Then (4) gives

$$\left|\frac{\dot{i}_z}{\dot{i}_0}\right| = e^{-\frac{(r_0-r)}{\delta}}$$

In Fig. 6.05b are plotted curves of $|\dot{i}_z/\dot{i}_0|$ using this formula, and comparisons are made with curves obtained from the exact formula, Eq. 6.04(8). This is done for two cases, $r_0/\delta = 2.39$ and $r_0/\delta = 7.55$. In the latter, the approximate distribution agrees well with the exact; in the former it does not. Thus if ratio of wire radius to δ is large, it seems that there should be little error in analyzing the wire from the results developed for plane solids. This point will be pursued later in impedance calculations.

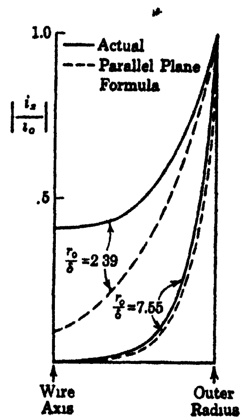


FIG. 6.05b. Actual and approximate (parallel plane formula) distribution in cylindrical wire.

Another approach would be to consider the trigonometric expressions which the Bessel functions approach for large arguments. When these expressions are valid (corresponding to arguments greater than about 10) it will be found that the trigonometric approximation is precisely the solution to the plane case.

Problem 6.05. Show that the trigonometric approximations for the Bessel functions of large arguments when substituted into the solutions of Art. 6.04 will yield precisely the solutions for the plane case of Art. 6.05.

6.06 Meaning of Internal Impedance

To extend the results of Arts. 6.04 and 6.05 to impedance calculations, it will be well to make first a convenient division of work in the impedance problem. To find total impedance of the circuit, the configuration of the actual conducting path must be known since this affects total inductance of the circuit. However, let us divide the inductance into two parts, one due to magnetic flux external to the wire, and one due to magnetic flux inside the wire itself. The first part is definitely dependent upon the shape of the conducting path. The second part is nearly independent of the external circuit, unless there are very sharp curvatures or other parts of the circuit so close to the wire that the

current distribution is seriously influenced by the proximity of those parts. The external portion of the inductance will be the subject of later study. To find internal inductance we shall then proceed with the analysis, using as the applied voltage to the wire the voltage at the surface of the wire. In an actual problem, this is really a resultant of the total applied voltage of the circuit and the induced voltage from magnetic flux outside the wire. That is, the drop due to external inductance has already been subtracted.

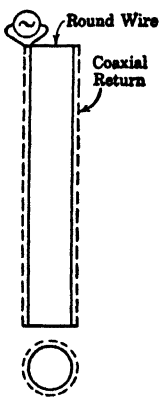


FIG. 6.06. Wire with coaxial return spaced an infinitesimal distance away.

A case in which the internal inductance has impressive meaning may be created by shrinking to zero the area available for magnetic flux external to the wire. In the limit such a case would result in a coaxial system as the space between inner and outer conductors vanishes. In the limit there would be no area external to the conductors in which magnetic flux might flow, and so no external inductance. There would be left only the internal inductance of the inner conductor and its return shield (Fig. 6.06).

To summarize: Internal impedance of a wire is defined as that part of the impedance of the circuit due to resistance drop in the wire and reactance drop from flux contained inside the wire.

Internal impedance of the wire may be obtained by dividing the total voltage *at the surface of the wire* by total current in the wire.

To obtain total impedance of an actual circuit, the external impedance, or reactance due to flux outside the wire, must be added to the internal impedance of the wire.

It should be especially noted that the above is the only intended meaning of the internal impedance to be calculated in following articles. It should not be interpreted as the impedance of a single wire alone without a return path. It should also be remembered that this division is strictly an approximation, and there may be circuit configurations, such as complicated coils, where it is not particularly useful.

Problem 6.06(a). Consider a single round wire of any finite radius in free space carrying a current I with no return path within a finite distance. Show that the magnetic flux surrounding the wire, and so the inductance per unit length, is infinite.

Problem 6.06(b). Show that the total magnetic flux and the inductance per unit length are infinite for a wire of infinitesimal size if it is carrying current I and its return path is removed from it by any finite distance.

6.07 Impedance of a Plane Conductor

The internal impedance for a unit width and unit length of the semi-infinite plane conductor of Art. 6.05 may now be found easily. Current density is given at any point in the conductor by Eq. 6.05(4). Total current flowing must be the integral of this from the surface to the infinite depth. For a unit width,

$$J_z = \int_0^\infty i_z\,dx = \int_0^\infty i_0 e^{-(1+j)\frac{x}{\delta}}\,dx = \frac{i_0\delta}{(1+j)} \qquad [1]$$

The electric intensity at the surface is given by the current density at the surface,

$$E_{z_0} = \frac{i_0}{\sigma}$$

Internal impedance, by the reasoning of Art. 6.06, is found from the quotient of this field and the current (1). For a unit length and unit width, or (as we shall often refer to it) per *unit square*

$$Z_s = \frac{E_{z_0}}{J_z} = \frac{(1+j)}{\sigma\delta} \qquad [2]$$

Define

$$Z_s = R_s + j\omega L_i$$

Then

$$R_s = \frac{1}{\sigma\delta} = \sqrt{\frac{\pi f \mu}{\sigma}}$$

$$\omega L_i = \frac{1}{\sigma\delta} = R_s \qquad [3]$$

The resistance and internal reactance of such a plane conductor are equal at any frequency. The internal impedance Z_s thus has a phase angle of 45°. Equation (3) gives another interpretation of depth of penetration δ, for this equation shows that the skin effect resistance of the semi-infinite plane conductor is exactly the same as the D-C resistance of a plane conductor of depth δ. That is, resistance of this conductor with exponential decrease in current density is exactly the same as though current were uniformly distributed over a depth δ.

R_s, the resistance of the plane conductor per unit square may be logically called a surface resistivity.

Values of depth of penetration and surface resistivity are given below for several commonly used materials. Curves of δ and R_s versus frequency are plotted in Fig. 6.05*a*. These values will be used extensively throughout the book.

	Conductivity mhos/meter σ	Permeability henrys/meter μ	Depth of Penetration meters δ	Surface Resistivity Ohms R_s
Silver	6.17×10^7	$4\pi \times 10^{-7}$	$\frac{0.0642}{\sqrt{f}}$	$2.52 \times 10^{-7}\sqrt{f}$
Copper	5.80×10^7	$4\pi \times 10^{-7}$	$\frac{0.0660}{\sqrt{f}}$	$2.61 \times 10^{-7}\sqrt{f}$
Aluminum	3.72×10^7	$4\pi \times 10^{-7}$	$\frac{0.0826}{\sqrt{f}}$	$3.26 \times 10^{-7}\sqrt{f}$
Brass	1.57×10^7	$4\pi \times 10^{-7}$	$\frac{0\ 127}{\sqrt{f}}$	$5.01 \times 10^{-7}\sqrt{f}$
Solder	$0\ 706 \times 10^7$	$4\pi \times 10^{-7}$	$\frac{0.185}{\sqrt{f}}$	$7.73 \times 10^{-7}\sqrt{f}$

Problem 6.07(*a*). Show that the magnetic field at the surface of the semi-infinite plane conductor, H_y, is equal in magnitude to J_z, current per unit width.

Problem 6.07(*b*). Show that for any general orientation of the uni-directional current in the semi-infinite plane conductor, magnetic field at the surface is related to linear current density $\bar{J}$ in magnitude and direction by

$$\bar{J} = \bar{n} \times \bar{H}$$

$\bar{n}$ is a unit vector normal to the surface and pointing away from the conductor.

6.08 Impedance of a Round Wire at Very High or Very Low Frequencies

Very High Frequency. To show the usefulness of impedance formulas for a semi-infinite plane solid, we shall obtain from them the internal impedance of a round wire at very high frequencies. It has already been shown that if the frequency is high enough, the curvature of the wire is unimportant. It may then be considered as a plane solid of practically infinite depth, and width equal to its circumference. Thus if Z_s, Eq. 6.07(2), is the internal impedance of the plane solid per unit square, $\frac{Z_s}{2\pi r_0}$ is the impedance for a width $2\pi r_0$ (the circumference). So, for a round wire of radius r_0 at very high frequencies,

$$Z_\infty = \frac{(1+j)}{2\pi r_0 \sigma \delta} = \frac{R_s(1+j)}{2\pi r_0}$$

or

$$R_\infty = (\omega L_i)_\infty = \frac{R_s}{2\pi r_0} \quad \text{ohms/meter} \qquad [1]$$

where

$$R_s = \frac{1}{\sigma\delta} = \sqrt{\frac{\pi f \mu}{\sigma}} \quad \text{ohms}$$

The notations Z_∞ and R_∞ are used to denote these very high-frequency formulas since they are exactly true only when frequency approaches infinity. The ratio of r_0/δ, however, serves as an indication of the frequencies for which they may be used, as we shall see from the exact analysis.

Very Low Frequency. For very low frequencies, the current has essentially a uniform distribution over the cross section (Fig. 6.04), and so the D-C resistance formula applies.

$$R_0 = \frac{1}{\pi r_0^2 \sigma} \quad \text{ohms/meter} \qquad [2]$$

By a method discussed in Art. 6.22, the internal inductance for a round wire with uniform current distribution can easily be found to be

$$(L_i)_0 = \frac{\mu}{8\pi} \quad \text{henrys/meter} \qquad [3]$$

The first correction term to resistance at moderately low frequencies may be obtained from series expansions of the exact results of the next article. This leads to

$$\frac{R}{R_0} = 1 + \frac{1}{48}\left(\frac{r_0}{\delta}\right)^4 \qquad [4]$$

The above equation is good for small values of r_0/δ and has an error of about 7 per cent at $r_0/\delta = 2$ (that is, for a radius twice the depth of penetration).

Problem 6.08(*a*). Show that the ratio of very high-frequency resistance to D-C resistance of a round conductor of radius r_0 and material with depth of penetration δ can be written

$$\frac{R_\infty}{R_0} = \frac{r_0}{2\delta}$$

Problem 6.08(*b*). By using the approximate formula 6.08(4), find the value of r_0/δ below which R differs from D-C resistance R_0 by less than 5 per cent. To what size wire does this correspond for

1. Copper at 10 kc/sec.
2. Copper at 1 mc/sec.
3. Brass at 1 mc/sec.

6.09 Impedance of Round Wires Generally

The internal impedance of the round wire at any frequency is found from total current in the wire and the electric intensity at the surface, according to the ideas of Art. 6.06. Total current may be obtained from an integration of current density, as for the plane conductor in Art. 6.07; however, it may also be found from the magnetic field at the surface, since the line integral of magnetic field around the outside of the wire must be equal to the total current in the wire.

$$\oint \bar{H} \cdot d\bar{l} = I$$

or

$$2\pi r_0 H_\phi \big|_{r=r_0} = I \tag{1}$$

Magnetic field is obtained from the electric field by Maxwell's equations.

$$\nabla \times \bar{E} = -j\omega\mu\bar{H} \tag{2}$$

For the round wire with the assumptions made in Art. 6.04, E_z and H_ϕ alone are present, and only r derivatives remain, so (2) is simply

$$H_\phi = \frac{1}{j\omega\mu}\frac{dE_z}{dr} \tag{3}$$

An expression for current density has already been obtained in Eq. 6.04(5). Electric field is related to this through the conductivity σ.

$$E_z = \frac{i_z}{\sigma} = \frac{i_0}{\sigma}\frac{J_0(Tr)}{J_0(Tr_0)} \tag{4}$$

By substituting in (3) and recalling that $T^2 = -j\omega\mu\sigma$

$$H_\phi = \frac{i_0 T}{j\omega\mu\sigma}\frac{J_0'(Tr)}{J_0(Tr_0)} = -\frac{i_0}{T}\frac{J_0'(Tr)}{J_0(Tr_0)}$$

$J_0'(Tr)$ denotes $\dfrac{d}{d(Tr)} J_0(Tr)$. From (1),

$$I = -\frac{2\pi r_0 i_0}{T}\frac{J_0'(Tr_0)}{J_0(Tr_0)} \tag{5}$$

The internal impedance per unit length is

$$Z_i = \frac{E_z\big|_{r=r_0}}{I} = -\frac{TJ_0(Tr_0)}{2\pi r_0\sigma J_0'(Tr_0)} \tag{6}$$

This is the complete expression for internal impedance of the wire. Z_i will be complex since T is complex.

$$T = \sqrt{-j\omega\mu\sigma} = j^{-\frac{1}{2}}\sqrt{2\pi f\mu\sigma} = \frac{\sqrt{2}\,j^{-\frac{1}{2}}}{\delta}$$

To separate into real and imaginary parts, use Eq. 6.04(6).

$$Ber\,v + jBei\,v = J_0(j^{-\frac{1}{2}}v)$$

Also let

$$Ber'\,v + jBei'\,v = \frac{d}{dv}\,(Ber\,v + jBei\,v)$$
$$= j^{-\frac{1}{2}}J_0'(j^{-\frac{1}{2}}v)$$

Then (6) may be written

$$Z_i = R + j\omega L_i = \frac{jR_s}{\sqrt{2}\,\pi r_0}\left[\frac{Ber\,q + jBei\,q}{Ber'\,q + jBei'\,q}\right]$$

where

$$R_s = \frac{1}{\sigma\delta} = \sqrt{\frac{\pi f\mu}{\sigma}} \qquad q = \frac{\sqrt{2}\,r_0}{\delta}$$

or

$$R = \frac{R_s}{\sqrt{2}\,\pi r_0}\left[\frac{Ber\,q\,Bei'\,q - Bei\,q\,Ber'\,q}{(Ber'\,q)^2 + (Bei'\,q)^2}\right] \quad \text{ohms/meter}$$
[7]
$$\omega L_i = \frac{R_s}{\sqrt{2}\,\pi r_0}\left[\frac{Ber\,q\,Ber'\,q + Bei\,q\,Bei'\,q}{(Ber'\,q)^2 + (Bei'\,q)^2}\right] \quad \text{ohms/meter}$$

These are the expressions for resistance and internal reactance of a round wire at any frequency in terms of the parameter q, which is $\sqrt{2}$ times the ratio of wire radius to depth of penetration. Curves giving the ratios of these quantities to the D-C and to the high-frequency values as functions of r_0/δ are plotted in Figs. 6.09a and 6.09b. A careful study of these will reveal the ranges of r_0/δ over which it is permissible to use the approximate formulas for resistance and reactance. For example, if a 10 per cent error can be

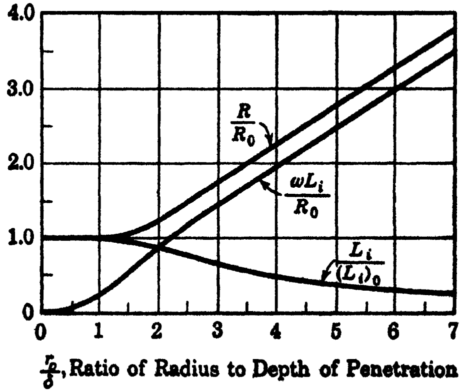


FIG. 6.09a. Solid wire skin effect quantities compared to D-C values.

tolerated, the high-frequency approximation for resistance, Eq. 6.08(1), may be used for $r_0/\delta > 5.5$; the high-frequency approximation for re-

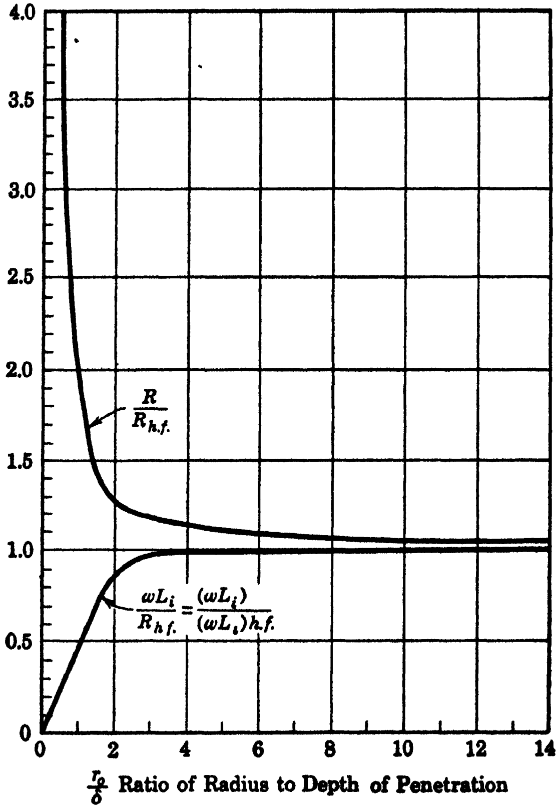


FIG. 6.09b. Solid wire skin effect quantities compared to values from high-frequency formulas.

actance, Eq. 6.08(1), may be used for $r_0/\delta > 2.2$. The D-C resistance formula Eq. 6.08(2) may be used for $r_0/\delta < 1.5$, and the D-C inductance formula Eq. 6.08(3) may be used for $r_0/\delta < 1.9$.

Problem 6.09(a). For small values of Tr_0 (low frequencies) the Bessel functions may be approximated by only a few terms of the series (Art. 3.18).

$$J_0(v) \cong 1 - \left(\frac{v}{2}\right)^2 + \frac{1}{4}\left(\frac{v}{2}\right)^4$$

$$J_0'(v) = -J_1(v) \cong -\left[\left(\frac{v}{2}\right) - \frac{\left(\frac{v}{2}\right)^3}{2!} + \frac{\left(\frac{v}{2}\right)^5}{2!\cdot 3!}\right]$$

Show that these values, substituted in Eq. 6.09(6), lead to the expressions for low-frequency resistance and inductance stated in Eq. 6.08(3) and Eq. 6.08(4).

Problem 6.09(*b*). For large values of Tr_0 (high frequencies) the Bessel functions may be approximated by the asymptotic forms of Art. 3.19. Show that these, substituted in Eq. 6.09(6), lead to the expressions for resistance and internal inductance at high frequencies obtained in Eq. 6.08(1).

Problem 6.09(*c*). From Figs. 6.09*a* and 6.09*b*, investigate the ranges of r_0/δ over which it is permissible to use the approximate formulas of Eqs. 6.08(1), 6.08(2), and 6.08(3) if the error must be less than 5 per cent.

6.10 Impedance of a Coated Conductor

Coated conductors appear in radio applications in the form of tinned copper wires, copper-plated iron or iron alloys in vacuum tube leads, silver-plated brass in resonant cavities, etc. Most often the problem is one of the following.

1. The coating material may be very thick compared with depth of penetration in that material. This requires no analysis since fields and currents in the coated metal are then negligible, and the impedance is governed only by the metal of the coating; the conductor is as good or as bad as a solid conductor of the coating material.

2. The coating may not be thick enough to prevent currents from flowing in the coated material, but penetration in both materials may be small compared with surface curvature so that the surfaces may be considered as planes, the coated material also being effectively infinite in depth. An analysis of this second case will follow.

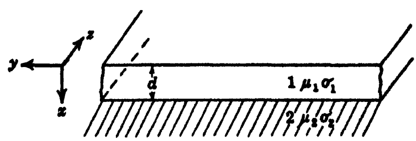


FIG. 6.10*a*. Conductor coated with another conductor.

In Fig. 6.10*a* is shown a plane solid material (conductivity σ_2, permeability μ_2) of effectively infinite depth coated with another material (conductivity σ_1, permeability μ_1) of thickness d. Solutions for the distribution equations must be found for both media and matched at the boundary between the two. The solution in either material is of the form of Eq. 6.05(2), but there can be no positive exponential term for the lower material since current density must become zero at infinite depth.

$$i_{z_2} = Ce^{-\tau_2 x}$$

$$\tau_2 = \frac{(1+j)}{\delta_2} = (1+j)\sqrt{\pi f \mu_2 \sigma_2} \qquad [1]$$

In the coating material both exponentials must be present, but it is

convenient to write the solution in terms of hyperbolic functions instead of the equivalent exponentials.

$$\dot{i}_{z_1} = A \sinh \tau_1 x + B \cosh \tau_1 x$$

$$\tau_1 = \frac{(1+j)}{\delta_1} = (1+j)\sqrt{\pi f \mu_1 \sigma_1} \qquad [2]$$

For both materials, as in Art. 6.07,

$$E_z = \frac{\dot{i}_z}{\sigma}$$

and

$$H_y = \frac{1}{j\omega\mu}\frac{dE_z}{dx} = \frac{\sigma}{\tau^2}\frac{dE_z}{dx}$$

Electric and magnetic fields in the two materials are then

$$E_{z_2} = \frac{C}{\sigma_2} e^{-\tau_2 x}; \quad E_{z_1} = \frac{1}{\sigma_1}[A \sinh \tau_1 x + B \cosh \tau_1 x]$$

$$H_{y_2} = -\frac{C}{\tau_2} e^{-\tau_2 x}; \quad H_{y_1} = \frac{1}{\tau_1}[A \cosh \tau_1 x + B \sinh \tau_1 x] \qquad [3]$$

The constants may be evaluated since tangential electric and magnetic fields are continuous across the boundary (Arts. 4.19 and 4.20).

$$E_{z_1} = E_{z_2}; \quad H_{y_1} = H_{y_2} \quad \text{at} \quad x = d$$

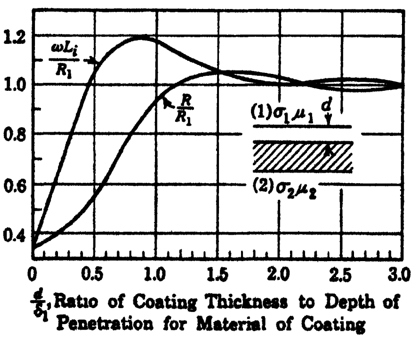


FIG. 6.10b. $\sqrt{\frac{\mu_2\sigma_1}{\mu_1\sigma_2}} = 0.34.$

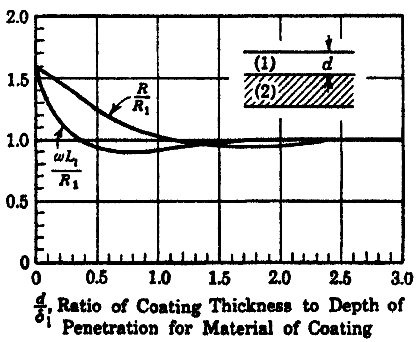


FIG. 6.10c. $\sqrt{\frac{\mu_2\sigma_1}{\mu_1\sigma_2}} = 1.6.$

Skin effect quantities for coated conductors.

Then

$$\frac{B}{A} = -\left[\frac{\sinh \tau_1 d + \dfrac{\tau_2\sigma_1}{\tau_1\sigma_2}\cosh \tau_1 d}{\cosh \tau_1 d + \dfrac{\tau_2\sigma_1}{\tau_1\sigma_2}\sinh \tau_1 d}\right] \qquad [4]$$

Total current in the two materials is obtainable from Prob. 6.07(b)

$$\bar{J} = \bar{n} \times \bar{H} \quad \text{or} \quad J_z = -H_y|_{x=0}$$

The impedance per square (per unit width and unit length) is

$$Z = \frac{E_z|_{x=0}}{J_z} = -\left.\frac{E_z}{H_y}\right|_{x=0} = -\frac{B}{A}\frac{\tau_1}{\sigma_1} \qquad [5]$$

From (4)

$$Z = \frac{\tau_1}{\sigma_1}\left[\frac{\sinh \tau_1 d + \dfrac{\tau_2\sigma_1}{\tau_1\sigma_2}\cosh \tau_1 d}{\cosh \tau_1 d + \dfrac{\tau_2\sigma_1}{\tau_1\sigma_2}\sinh \tau_1 d}\right]$$

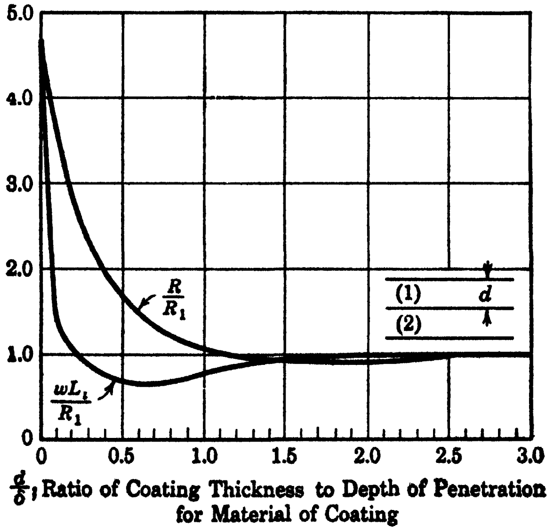


FIG. 6.10d. Skin effect quantities for coated conductor: $\sqrt{\dfrac{\mu_2\sigma_1}{\mu_1\sigma_2}} = 5.$

But, by the definition of R_s, Eq. 6.07(3), this may be written

$$\frac{Z}{R_{s_1}} = (1 + j)\left[\frac{\sinh \tau_1 d + \dfrac{R_{s_2}}{R_{s_1}} \cosh \tau_1 d}{\cosh \tau_1 d + \dfrac{R_{s_2}}{R_{s_1}} \sinh \tau_1 d}\right] \qquad [6]$$

Curves of ratio of resistance and reactance of this conductor to resistance of a conductor made entirely of the coating material are shown in Fig. 6.10*b*, *c*, and *d*. The curves are plotted as functions of d/δ_1, which is the ratio of coating thickness to depth of penetration for material of the coating (given by Fig. 6.05*a*). The several values of ratio R_{s_2}/R_{s_1} selected correspond approximately to solder on copper, silver on brass, and copper on iron. It is seen that in each case the composite conductor becomes roughly as good, or as bad, as though the coating were of infinite depth, when the coating reaches a thickness equal to δ_1, depth of penetration for the material of the coating.

6.11 Impedance of Tubular Conductors

A study of the current distribution in a solid round wire shows that at the higher frequencies the inner part of the conducting material plays little part in the conduction. There should consequently be little difference in impedance under such skin effect conditions between a solid round wire and a hollow tubular conductor of the same outer diameter. Certainly this is true when the wall thickness is very large compared with the depth of penetration of the conducting material, and we would use the same high-frequency equation, Eq. 6.08(1), that was developed for a round wire. If this criterion is not satisfied, the finite wall thickness must be taken into account. An exact solution may be carried through in terms of Bessel functions, but many times the wall thickness is small enough compared with wire radius so that the analysis of a flat plane conductor of finite thickness may be applied well enough. The result for this problem may be lazily found at once by setting conductivity of the lower material equal to zero in the result for the composite conductor of Art. 6.10. That is, in Eq. 6.10(6), set $R_{s_2} = \infty$. Then, the surface impedance per square

$$Z = R + j\omega L_i = (1 + j)R_s \frac{\cosh \tau d}{\sinh \tau d} \qquad [1]$$

$$= (1 + j)R_s \coth\left[\frac{d}{\delta}(1 + j)\right] \qquad [2]$$

$$\frac{R}{R_s} = \frac{\sinh\left(\dfrac{2d}{\delta}\right) + \sin\left(\dfrac{2d}{\delta}\right)}{\cosh\left(\dfrac{2d}{\delta}\right) - \cos\left(\dfrac{2d}{\delta}\right)} \qquad [3]$$

For a tubular conductor satisfying the conditions of wall thickness small compared with radius of tube these results may be used directly to give impedance per unit length by dividing by the circumference. In applying these results to tubular conductors in which the exciting sources are on the inside, as, for example, the outer conductor of a coaxial line, it may not be clear whether inner or outer radius of the tube should be used in the formula. Certainly the formula is not exact with either, and the assumption is that there is little difference between the two radii. However, it seems to make more sense to use inner radius for inner exciting sources since fields are applied from the inside, and (at high frequencies) currents concentrate on the inner surface. Thus, for the tubular conductor

$$R = \frac{R_s}{2\pi r}\left[\frac{\sinh\left(\dfrac{2d}{\delta}\right) + \sin\left(\dfrac{2d}{\delta}\right)}{\cosh\left(\dfrac{2d}{\delta}\right) - \cos\left(\dfrac{2d}{\delta}\right)}\right] \qquad [4]$$

$$\omega L_i = \frac{R_s}{2\pi r}\left[\frac{\sinh\left(\dfrac{2d}{\delta}\right) - \sin\left(\dfrac{2d}{\delta}\right)}{\cosh\left(\dfrac{2d}{\delta}\right) - \cos\left(\dfrac{2d}{\delta}\right)}\right] \qquad [5]$$

where r = outer radius if fields are applied along outside of tube.
r = inner radius if fields are applied along inside of tube.

The high-frequency resistance of the tubular conductor is merely

$$R_{hf} = (\omega L_i)_{hf} = \frac{R_s}{2\pi r}$$

Curves of resistance and internal reactance ratios to high-frequency resistance are given in Fig. 6.11*b* and to D-C resistance are given in Fig. 6.11*a*.

Problem 6.11 (*a*). Show that for a tubular conductor of outer radius r_0, inner radius r_i with voltage applied from the outside, the exact expressions for skin effect resistance and reactance are

$$R + j(\omega L_i) = \frac{j^{-1/2}\sqrt{2}R_s}{2\pi r_0}\left[\frac{J_0(Tr_0)H_0^{(1)\prime}(Tr_i) - J_0'(Tr_i)H_0^{(1)}(Tr_0)}{J_0'(Tr_0)H_0^{(1)\prime}(Tr_i) - J_0'(Tr_i)H_0^{(1)\prime}(Tr_0)}\right]$$

T as in Eq. 6.04(2).

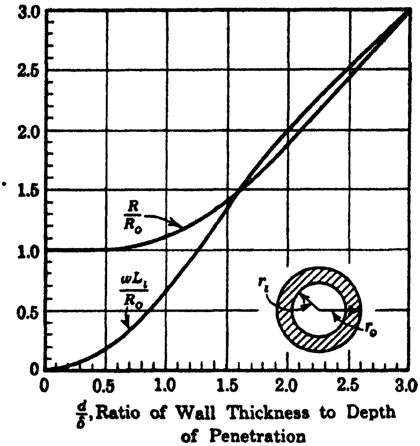


FIG. 6.11*a*. Thin-walled tubular conductor. Skin effect quantities compared with D-C values.

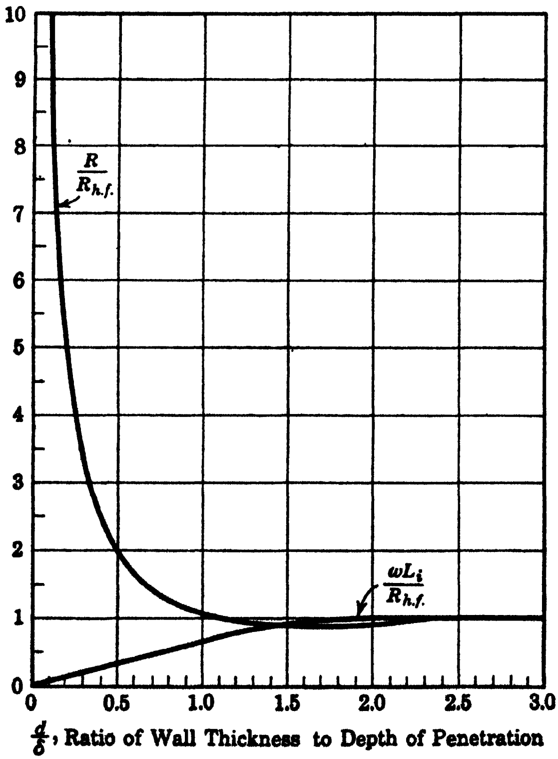


FIG. 6.11*b*. Thin-walled tubular conductor. Skin effect quantities compared with values from high-frequency formulas.

Problem 6.11(*b*). For a case with $r_0/\delta = 1.25$ and $r_i/\delta = 1.0$, calculate the skin effect resistance from the result of the preceding problem and compare with that calculated from the approximate expression, Eq. 6.11(4).

Problem 6.11(*c*). Show that the result of Prob. 6.11(*a*) may be used for a tubular conductor with voltage applied at the inner radius, if r_0 and r_i are interchanged.

6.12 Similitude Relations for Skin Effect

Very often it is desirable to compare two conductor sizes, or two different metals, or coatings, etc., as to their high-frequency impedance. From the relations of the preceding articles, or directly from Maxwell's equations, certain general similitude relations may be proved.

1. If two systems are geometrically similar and of the same material, the current distributions will be similar, and current densities will be equal in amplitude at similar points if the applied voltage to the small system is $1/K$ in magnitude and K^2 in frequency that of the large system, where K is the ratio of linear dimensions of the large to small system.

2. The impedance of the small system will be K times that of the large system under the conditions of statement 1.

3. For the distribution of currents and fields to be similar in general, two equations must be satisfied.[2]

$$\mu_L \epsilon_L K^2 \omega_L^2 = \mu_S \epsilon_S \omega_S^2$$

$$\mu_L \sigma_L K^2 \omega_L = \mu_S \sigma_S \omega_S$$

where the subscripts L and S refer to large and small systems respectively.

CALCULATION OF INDUCTANCE

6.13 Definition of Mutual Inductance

A change in the current of a circuit causes changing magnetic effects in the region of other neighboring circuits. These changing magnetic effects induce electric fields in the region of any neighboring circuits and consequently voltages around these circuits. For the two circuits pictured in Fig. 6.13 a change of the current in circuit 1 induces a voltage in circuit 2. In the previous chapter we showed how this phenomenon leads to a concept of mutual inductance. Mutual inductance between circuit 1 and circuit 2 is defined in conventional circuit theory as the

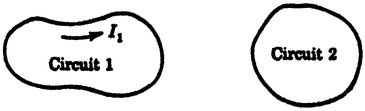


FIG. 6.13.

[2] J. A. Stratton, "Electromagnetic Theory," McGraw-Hill, 1941, p. 489.

coefficient relating the induced voltage to the changing current.

$$V_{21} = -M_{21} \frac{dI_1}{dt} \qquad [1]$$

6.14 Mutual Inductance from Vector Potential

The electric field at any point induced by changing magnetic effects due to the current in circuit 1 (Fig. 6.13) is

$$\bar{E}_1 = -\mu \frac{\partial \bar{A}_1}{\partial t} \qquad [1]$$

$\bar{A}_1$ is given in terms of the current in circuit 1 by

$$\bar{A}_1 = \int_V \frac{[\bar{\imath}_1]\, dV}{4\pi r} \qquad [2]$$

Making the assumptions that retardation, displacement currents, and distributed effects are all negligible, $\bar{A}_1$ will be proportional at every point to total current I_1. Hence,

$$V_{21} = -\oint \mu \frac{\partial \bar{A}_1}{\partial t} \cdot d\bar{l}_2 = -\frac{\partial I_1}{\partial t} \frac{\oint \mu \bar{A}_1 \cdot d\bar{l}_2}{I_1} \qquad [3]$$

Comparing with Eq. 6.13(1) and recognizing the equivalence of $\partial I_1/\partial t$ with dI_1/dt (Art. 5.04) identifies mutual inductance as

$$M_{21} = \frac{\oint \mu \bar{A}_1 \cdot d\bar{l}_2}{I_1} \qquad [4]$$

This expression is completely analogous to that which was identified as self inductance in the previous chapter.

The use of vector potentials directly is especially helpful in the calculation or estimate of mutual inductances for current paths having straight line portions. Consider, for instance, the rectangular circuits of Fig. 6.14a. Examination of (2) shows that $\bar{A}$ can have only the direction of the current producing it; induced field

$$\bar{E} = -\mu \frac{\partial \bar{A}}{\partial t} = -j\omega\mu \bar{A}$$

has the direction of $\bar{A}$. Consequently in the system of Fig. 6.14a there can be no contribution to voltage in the sides a_2 and b_2 from current in the sides c_1 and d_1, nor any contribution in the sides c_2 and d_2 from current in the sides a_1 and b_1. The picture obtained from visualizing these

directive relations is frequently valuable where the circuit configuration is quite complex. Often a quick estimate of coupling may be obtained by replacing portions of current paths by straight line sections, estimating the coupling from these.

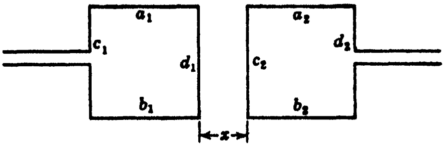


FIG. 6.14*a*. Two square coupling loops.

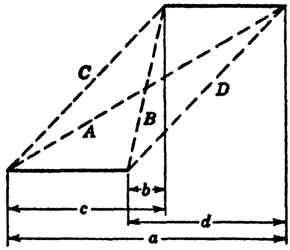


FIG. 6.14*b*. Parallel current elements displaced from one another.

Problem 6.14(*a*). By integration of Eq. 6.14(4), show that the contribution to mutual inductance from two parallel line segments displaced as shown in Fig. 6.14*b* is

$$M = \frac{\mu}{4\pi}\left[\ln\left\{\frac{(A+a)^a(B+b)^b}{(C+c)^c(D+d)^d}\right\} + (C+D) - (A+B)\right]$$

Problem 6.14(*b*). Apply the above result to the calculation of mutual inductance between two square loops used for coupling between open-wire transmission lines as shown in Fig. 6.14*a*. The length of each side is 0.03 meter; the separation x is 0.01 meter. Assume that the gaps at which the lines enter are small enough to be ignored.

6.15 Neumann's Form for Mutual Inductance

When field distribution in the region of the second circuit due to current in the first is essentially independent of the manner in which current is distributed over the cross sections of conductors in the first circuit, this field may be calculated by assuming all the current concentrated in a thin filament. The integral expression for $\bar{A}_1$ in terms of the total current I_1 flowing in circuit 1 is then

$$\bar{A}_1 = \oint \frac{\bar{I}_1\, dl_1}{4\pi r} = \oint \frac{I_1\, \overline{dl}_1}{4\pi r} \qquad [1]$$

This value of $\bar{A}_1$ may be substituted in Eq. 6.14(4) for mutual inductance

$$M_{21} = \mu \oint\oint \frac{\overline{dl}_1 \cdot \overline{dl}_2}{4\pi r} \qquad [2]$$

This is Neumann's form for obtaining mutual inductance between circuits. It is the same expression as Eq. 6.14(4), except that it assumes

that $\bar{A}_1$ in the region of circuit 2 may be calculated well enough by concentrating the current in circuit 1 in a thin filament. Its use will be demonstrated by the example of the next article.

6.16 Mutual Inductance between Two Coaxial Circular Loops

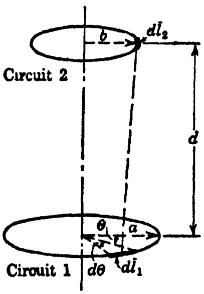


FIG. 6.16. Two coaxial circular loops.

Neumann's form, Eq. 6.15(2), will be used to obtain the value of mutual inductance between two coaxial loops as pictured in Fig. 6.16. If $\bar{dl}_1$ is any element of circuit 1, and $\bar{dl}_2$ is any element of circuit 2,

$$\bar{dl}_1 \cdot \bar{dl}_2 = dl_2\, a\, d\theta \cos\theta$$

$$r = \sqrt{d^2 + (a \sin\theta)^2 + (a\cos\theta - b)^2}$$

By substituting $\theta = \pi - 2\phi$ and

$$k^2 = \frac{4ab}{d^2 + (a+b)^2}$$

the integral then will be found to become

$$M = \mu\sqrt{ab}\, k \int_0^{\pi/2} \frac{(2\sin^2\phi - 1)\, d\phi}{\sqrt{1 - k^2\sin^2\phi}}$$

This can be broken into two integrals.

$$M = \mu\sqrt{ab}\left[\left(\frac{2}{k} - k\right)\int_0^{\pi/2} \frac{d\phi}{\sqrt{1 - k^2\sin^2\phi}} - \frac{2}{k}\int_0^{\pi/2} \sqrt{1 - k^2\sin^2\phi}\, d\phi\right]$$

$$= \mu\sqrt{ab}\left[\left(\frac{2}{k} - k\right)K(k) - \frac{2}{k}E(k)\right]$$

where

$$E(k) = \int_0^{\pi/2} \sqrt{1 - k^2\sin^2\phi}\, d\phi \qquad [1]$$

$$K(k) = \int_0^{\pi/2} \frac{d\phi}{\sqrt{1 - k^2\sin^2\phi}} \qquad [2]$$

The definite integrals (1) and (2) are tabulated in tables[3] as functions of k and are called complete elliptic integrals of the first and second kinds

[3] Dwight, "Tables of Integrals," Macmillan, 1934.

respectively. Thus

$$M = \mu\sqrt{d^2 + (a+b)^2}\left[\left(1 - \frac{k^2}{2}\right)K(k) - E(k)\right] \qquad [3]$$

where

$$k = 2\sqrt{\frac{ab}{d^2 + (a+b)^2}}$$

6.17 Mutual Inductance from Flux Linkages

The expression for mutual inductance may also be written in the familiar form based on the amount of flux from current in circuit 1 which links circuit 2. From Stokes' theorem, Art. 2.26,

$$\oint \mu\bar{A}_1 \cdot d\bar{l}_2 = \int_S \mu[\nabla \times \bar{A}_1] \cdot d\bar{S}_2$$

But

$$\nabla \times \bar{A}_1 = \bar{H}_1$$

So Eq. 6.14(4) may be written

$$M_{21} = \frac{\displaystyle\int_S \mu\bar{H}_1 \cdot d\bar{S}_2}{I_1} \qquad [1]$$

The surface integral $\int_S \mu\bar{H}_1 \cdot d\bar{S}_2$ is identified as the flux from circuit 1 enclosed by circuit 2. This flux linkage definition is perhaps the most familiar definition for inductance. It is identical with the definition in terms of vector potential for a circuit without discontinuities and differs very little if the discontinuity is small. Although the two forms are equivalent, one may be more convenient than the other under certain conditions.

If inductance is to be calculated from flux enclosed, any of the appropriate methods studied in Chapters 2 and 3 may be used for calculating fields in the region of circuit 2. For instance, suppose this method were applied to the two coaxial circular conductors of Fig. 6.16. H on the axis might then be written by application of Ampère's law which, except for the present use of mks units, was done in Art. 3.29.

$$H_z\Big|_{\text{axis}} = \frac{a^2 I_1}{2[z^2 + a^2]^{3/2}} = \frac{I_1}{2a}\left[1 - \frac{3}{2}\left(\frac{z}{a}\right)^2 + \cdots\right] \qquad [2]$$

For any point off the axis, it may be written in spherical coordinates as a

series of spherical harmonics (Art. 3.28)

$$H_z = \frac{I_1}{2a}\left[1 - \frac{3}{2}\left(\frac{r}{a}\right)^2 P_2(\cos\theta) + \cdots\right] \tag{3}$$

This expression may then be integrated over the region of circuit 2. However, if d/a is much greater than b/a (Fig. 6.16), or if b/a is small with any value of d, a study of the series terms of (3) shows that H_z does not vary much over the region of 2. A good approximation to inductance will result in these cases by assuming that H_z is constant over the area of the second loop. Then,

$$M_{21} \cong \frac{\mu\pi\, a^2b^2}{2(d^2 + a^2)^{3/2}} \text{ henrys} \tag{4}$$

6.18 Mutual Inductance by Reciprocity

A study of energy considerations would show that voltage induced in one circuit by a changing current in a second circuit is the same as the voltage induced in the second by the same rate of change of current in the first. That is,

$$M_{21} = M_{12}$$

For example, if it is desired to calculate voltage induced in the large circular loop 1 owing to a change in current in the small loop 2 (Fig. 6.16), it is not necessary to calculate the flux enclosed by loop 1 from the current in loop 2. The mutual inductance calculated in the previous article for flux linking circuit 2 due to current in circuit 1 may be used directly.

6.19 Neumann's Form Applied to Self Inductance

Self inductance may be considered merely as a special case of mutual inductance in that the circuit around which we wish to evaluate the voltage induced from changing magnetic effects is the circuit in which the current flows which produces these magnetic effects. The basis for calculation of self inductance of course rests ultimately on the concepts introduced in Chapter 5. These concepts certainly permit the notion of self inductance as a special case of mutual inductance.

It is apparent that certain of the forms derived for mutual inductances would not, however, be suitable for calculations of self inductance. For example, Neumann's form (Art. 6.15) if applied directly to the calculation of self inductance, using the same line path for both integrations, will always give an answer of infinity, since this corresponds to a calculation of the inductance of a circuit with an infinitesimal wire. This is not an incorrect result. The inductance of any infinitesimal wire is indeed infinite (Prob. 6.06b).

For a complete computation of self inductance, it is then necessary to take into account the finite cross section of conducting paths, and the actual distribution of currents over these cross sections. This part of the problem has been covered in early sections of the chapter. What we want to do now is to compute the external self inductance only, so we shall be interested in methods other than Neumann's for such calculations.

6.20 Self Inductance by Selected Mutual Inductance or Flux Linkage Method

In Art. 6.06, it was shown that for any circuit the induced voltage from changing magnetic effects, and hence the inductance, may be broken into two components, one due to flux inside the wire and the other from flux outside the wire. The internal inductance, or contribution from flux inside the wire, has been found for different shapes of conductors in Arts. 6.07 to 6.11. There remains the contribution from flux outside the wire.

Figure 6.20 shows that the induced voltage about a path taken along the surface of the conductor along the inner contour of the loop is obtained from flux enclosed by that path. This represents the desired contribution to inductance from flux external to the wire. At a point P some distance from the wire, the field is much the same for a given current in the conductor no matter how that current may be distributed over the conductor's cross section. For such a point it is nearly correct to calculate field intensity, assuming all current concentrated at the center of the conductor. Similarly, at point Q near the wire the field is $I/2\pi r$ (see Art. 2.36), where I is the total current in the conductor, regardless of how current is distributed over the cross section of the conductor, *provided other portions of the conducting path are not near enough to disturb the circular symmetry.* Field near the wire is then also very nearly the same as though current were concentrated at the center. We conclude that field at any point inside the loop may be calculated approximately by assuming a current concentrated along the axis of the wire. The problem of finding contribution to self inductance from external flux is then very nearly that of finding the mutual inductance between a line current along the axis of the wire and a line circuit

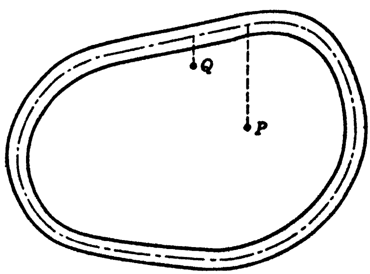


FIG. 6.20. Arbitrary circuit formed of a conductor of circular cross section.

selected around the inner surface of the loop. This way of looking at the problem shows that we can use mutual inductance formulas for calculating external self inductance. If desired, of course, we can simply compute the external flux linked by the wire and regard external inductance as the flux linkages per unit current. Since it is understood that this applies only to external inductance, there will be no need to consider any difficulties previously mentioned as attached to Neumann's form.

6.21 Self Inductance of a Circular Loop

As an example of the method described in Art. 6.20, which we have called the selected mutual inductance method, let us find the self inductance of a circular loop of wire. The wire radius is a, and the loop radius is r. The contribution to inductance from external flux, given by the mutual inductance between two concentric circles of radii r and $(r - a)$ may be obtained from Eq. 6.16(3).

$$L_0 = \mu(2r - a)\left[\left(1 - \frac{k^2}{2}\right)K(k) - E(k)\right] \qquad [1]$$

$$k^2 = \frac{4r(r - a)}{(2r - a)^2}$$

$$r = \text{radius of loop}$$

$$a = \text{radius of wire}$$

$K(k)$ and $E(k)$ are complete elliptic integrals of first and second kinds as defined by Eqs. 6.16(1) and 6.16(2). If a/r is very small, k is nearly unity, and K and E may be approximated by

$$K(k) \cong \ln\left(\frac{4}{\sqrt{1 - k^2}}\right)$$

$$E(k) \cong 1$$

so

$$L_0 \cong r\mu\left[\ln\left(\frac{8r}{a}\right) - 2\right] \text{ henrys} \qquad [2]$$

To find total L, values of internal inductance, as listed in Arts. 6.08 and 6.09, must be added.

6.22 Self Inductance by Energy Integrals

When retardation is neglected, the inductive impedance of a circuit does not cause any energy loss. Energy leaves the source to be distributed in space in the magnetic field. For an A-C source this stored

energy in the field is returned to the source later in the cycle. It is possible to relate the instantaneous stored energy in the magnetic field (Art. 4.24) to the instantaneous current in the conductor of the circuit which gives rise to the field. Thus

$$\tfrac{1}{2}LI^2 = \int_V \frac{\mu H^2}{2}\,dV \qquad [1]$$

It is often convenient to make use of this expression to compute the inductance of a circuit. Such a method is a good one, of course, if the configuration happens to be one for which the distribution function for H is easily evaluated and integrated.

In general, it may be thought that if the value of H is known everywhere, one cannot do better in choosing a method for calculating inductance than the flux linkage approach. This is probably true for the external inductance of a circuit. When, as in internal inductance computations, the flux which contributes to inductance exists in a current-carrying region, there are partial linkages to consider, and the flux linkage method becomes considerably more complex than the stored energy method.

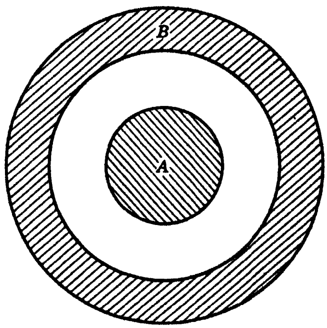


FIG. 6.22. Cross section of coaxial circuit.

Consider for example the long cylindrical conductor A with its coaxial return conductor B (shown in cross section in Fig. 6.22), constituting a complete circuit at a frequency low enough so that the current is distributed uniformly over the conductor cross sections. From symmetry, the flux lines are known to be circles about the axis. From Biot and Savart's law directly we can relate the strength of H everywhere to the current. It is found in this way that the flux is zero external to the outer conductor since the total current enclosed is zero. H is known in the three regions of interest: (1) inside the inner conductor (which yields part of the internal inductance), (2) inside the outer conductor (which gives the remainder of the internal inductance), (3) in the space between conductors (which gives the external inductance).

Problem 6.22(a). Find the internal and the total inductance per unit length of the circuit indicated in Fig. 6.22, first by the stored energy method and then by the flux linkage method.

Problem 6.22(*b*). Show that the internal inductance per unit length of a solid wire carrying low-frequency (thus uniformly distributed) current is (Art. 6.08):

$$(L_i)_0 = \frac{\mu}{8\pi} \quad \text{henrys/meter}$$

6.23 Inductance of Practical Coils

A study of the inductance of coils at low frequencies involves no new concepts but only new troubles because of the complications in geometry.

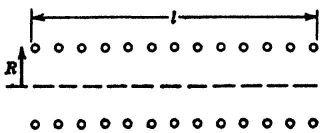


FIG. 6.23*a*. Longitudinal section of long solenoid.

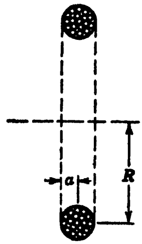


FIG. 6.23*b*. Cross section of a coil.

rtain cases are simple enough for calculation by a straightforward plication of previously outlined methods. For example, the inductce of the long solenoid of Fig. 6.23*a*, neglecting end effects, is found in s way to be approximately

$$L_0 = \pi\mu R^2 n^2[\sqrt{l^2 + R^2} - R] \quad \text{henrys} \qquad [1]$$

ere the symbols are as shown in the figure and n is the number of ns per meter.

For the other extreme, for coils whose cross sections are very small npared with radius (Fig. 6.23*b*) the equation for inductance of a g may be used to give approximate results, introducing only the nber of turns in the coil. From Eq. 6.21(2)

$$L_0 = RN^2\mu\left[\ln\left(\frac{8R}{a}\right) - 2\right] \qquad [2]$$

Many practical coil shapes will be such that either of these formulas would provide poor approximations. The coil configuration in these cases often makes integration so difficult that it is desirable to approximate inductance from empirical and semi-empirical formulas, such as those compiled by the Bureau of Standards. These formulas[4] are assumed to be readily available and so will not be repeated here.

[4] Circular 74, "Radio Instruments and Measurements," National Bureau of Standards.

At higher frequencies the problem becomes more complex. When turns are relatively close together, the assumption made previously in calculating internal impedance (other portions of the circuit so far away that circular symmetry of current in the wire is not disturbed) certainly does not apply. Current elements in neighboring turns will be near enough to produce nearly as much effect upon current distribution in a given turn as the current in that turn itself. Values of skin effect resistance and internal inductance are then not as previously calculated. External inductance may also be different since changes in external fields result when current loses its symmetrical distribution with respect to the wire axis.

In coils used in radio engineering, the relation between energy stored in the field to that dissipated in ohmic losses is often important, so that the ratio $\omega L/R$, the Q of the coil, is used to compare different coils. This factor could be calculated by the methods used earlier in this chapter, but for coils at high frequencies, the changes in current distribution just discussed must be taken into account. In general, the action from currents in neighboring turns always tends to concentrate current in a smaller portion of the wire cross section so that the actual resistance is always higher than that calculated, assuming symmetrical distributions, hence Q is always lower. In making calculations of the Q of coils, it must be kept in mind that losses in the insulating forms at high frequencies may still further increase the effective resistance and so decrease Q. Some useful guides for estimating Q of coils at high frequencies are given in a recent paper[5] which also contains a bibliography covering this field.

SELF AND MUTUAL CAPACITANCES

6.24 Definitions of Self and Mutual Capacitances

According to usual concepts, capacitance between two conductors is defined as the charge on either conductor divided by the potential difference between them. It was this kind of expression that was arrived at in Chapter 5, the assumption having been made that charges were concentrated at the circuit discontinuities which were called condensers or capacitors. Of course, the concept of capacitance was pictured as more general, often involving several conductors. It often happens in practical problems that more than two conductors influence the electric field in their vicinity, and the circuits which connect adjacent conductors are thus capacitively coupled. Wherever the more general definition of

[5] H. A. Wheeler, "Formulas for the Skin Effect," *Proc. I.R.E.*, **30**, September, 1942.

capacitance applies, the relation between the charges on conductors and the differences of potential between them is not so simple as in the two-conductor case where one has only to say that the charge is equal and opposite on the two conductors and proportional to the voltage between them.

For quantitative analysis as well as for the purpose of obtaining a better picture of this type of phenomenon, we shall need merely to systematize the concepts and the equations with which we have already dealt. The ultimate desire in circuit analysis is the relation between voltage drops and currents in a network of impedances. The first step is to obtain certain basic relations which will yield the charges on various electrostatically coupled conductors in terms of the differences of potential applied between them.

Given a system of several conductors each of which may be at any potential, consider first one of these. If a charge is placed on this conductor with all others charge free, the potential of the conductor might be found from

$$\Phi = \int_V \frac{\rho\, dV}{4\pi\epsilon r} \qquad [1]$$

Since this is a linear relation and there is no charge but that on the body itself, the potential finally calculated must be proportional to total charge on the body.

$$\Phi_{11} = p_{11}Q_1$$

Suppose next a second conductor is allowed to acquire a charge. Equation (1) could again be applied to calculate the added potential of body 1 due to the charge on 2, and this would be proportional to the charge on 2 (as $p_{21}Q_2$). Because of the linear character of causes and effects, superposition is allowable, and total potential of body 1 is now

$$\Phi_1 = p_{11}Q_1 + p_{12}Q_2$$

The process may be repeated as charges are placed on all the bodies in turn, so finally an entire set of equations may be written

$$\begin{aligned} \Phi_1 &= p_{11}Q_1 + p_{21}Q_2 + \cdots p_{n1}Q_n \\ \Phi_2 &= p_{12}Q_1 + p_{22}Q_2 + \cdots p_{n2}Q_n \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ \Phi_n &= p_{1n}Q_1 + p_{2n}Q_2 + \cdots p_{nn}Q_n \end{aligned} \qquad [2]$$

The coefficients p are called the coefficients of potential.

The linear set of equations (2) may be solved for any of the charges.

The results may be written in the form

$$
\begin{aligned}
Q_1 &= C_{11}\Phi_1 + C_{21}\Phi_2 + \cdots C_{n1}\Phi_n \\
Q_2 &= C_{12}\Phi_2 + C_{22}\Phi_2 + \cdots C_{n2}\Phi_n \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
Q_n &= C_{1n}\Phi_1 + C_{2n}\Phi_2 + \cdots C_{nn}\Phi_n
\end{aligned}
\qquad [3]
$$

The new coefficients of proportionality C may be called coefficients of capacity. That of the form C_{rr} represents the ratio of charge on the rth conductor to potential on that conductor with all other conductors grounded. It may be called the self capacity of the rth conductor. That of the form C_{rs} represents the ratio of charge induced on conductor s to potential on body r, all conductors but r grounded. This may then be called the mutual capacitance between r and s, although it is more often called the coefficient of induction.

6.25 Properties of the Coefficients of Capacity and Potential

A most important relation amongst the coefficients is that

$$C_{rs} = C_{sr} \qquad [1]$$

Green's reciprocation theorem[6] shows that this must be so, It is also possible to show that

$p_{rs} = p_{sr}$.

All p's are positive or zero.

C_{rr} is positive or zero.

$C_{rs}(r \neq s)$ is negative or zero.

The sum $C_{r1} + C_{r2} + C_{r3} + \cdots$ is zero or positive.

Problem 6.25. Show that the system of Eqs. 6.24(3), when applied to a simple capacitor consisting of two electrodes, reduces to the usual

$$Q = CV$$

where Q is the charge on either electrode, V is the voltage between them, and C is a constant. Also evaluate C in terms of the self and mutual capacitances.

6.26 Electrostatic Shielding

The relation between coefficients is brought out clearly by a study of one conductor perfectly shielded by a second. Consider such a conductor in the spherical condenser of Fig. 6.26a, with an external conducting body in the vicinity. We know at once that a change in the potential of 3 can in no way influence the charge on 1 because of the

[6] Smythe, "Static and Dynamic Electricity," McGraw-Hill, 1939, Chapter II.

completely surrounding grounded conductor 2. Thus $C_{13} = 0$. Also, if all conductors but 1 are grounded, the induced charge on 2 is only $-Q_1$, so the equations show that $C_{12} = -C_{11}$. This equal and opposite character of the self and mutual coefficients is a criterion of perfect shielding.

It is not necessary that a shield completely enclose a conductor to shield it from a third conductor as above. Thus in Fig. 6.26*b*, the presence of the grounded plane conductor 2 causes less charge to be induced on 3 when it is grounded and a given potential is placed on 1

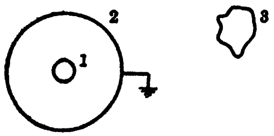


FIG. 6.26*a*. Electrostatic shielding by a grounded sphere.

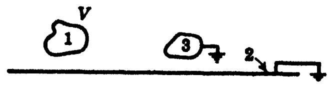


FIG. 6.26*b*. Partial shielding by a grounded conducting plane.

than when the plane is absent. This can easily be seen by constructing the images of 1 and 3 below the plane to replace the effect of the plane. Then it is evident that the new image conductors result in an induction effect on 3 which opposes the original effect of 1.

The shielding effect of body 2 in Figs. 6.26*a* and 6.26*b* would quickly disappear if the shield were ungrounded and insulated instead. The addition of a new insulated conductor generally causes an *increase* rather than a decrease in capacitance between two conductors, although this is not always true. If, for example, the two original conductors are concentric spheres, the addition of a third insulated concentric sphere between them will have no shielding or other effect on the field between conductors. The insulated sphere being an equipotential surface of the original field simply assumes that proper potential, induced charge appears in equal amounts on both its inside and outside surfaces, and the field is, in effect, transmitted from innermost to outermost sphere. Obviously, the addition of an insulated conductor to any field in the form of an equipotential surface of that field will always result in no shielding effect and no increase or decrease of capacity coupling between the original conductors of the field.

That the coupling is most always increased if the added conductor is insulated is illustrated by the case of Fig. 6.26*b*. If the plane is insulated, rather than grounded, the coupling between 1 and 3 will be found to be increased. In fact, suppose 1 and 3 are widely separated so that their coupling in the absence of the plane (or with the plane grounded) would approach zero. Then, if the plane is insulated, the coupling can

be shown to approach a constant value, depending upon the distance of 1 and 3 to the plane 2. Thus the coupling-increase ratio (insulated plane to no plane) approaches infinity as the bodies 1 and 3 are increasingly separated.

From the various foregoing examples it is seen that if a new electrode is introduced between or near two others for shielding purposes, it is very important that it be truly grounded. It often happens that such electrodes, although grounded for D-C, may be effectively insulated or floating at radio frequency because of impedance in the grounding leads. In such cases the new electrodes do not accomplish their shielding purpose. In fact, as indicated above, they may actually increase capacity coupling.

7

PROPAGATION AND REFLECTION OF ELECTROMAGNETIC WAVES

7.01 Introduction

As soon as Maxwell's equations are applied to physical systems, as in the circuits of Chapter 5, it is observed that in general, effects from all currents and charges are characterized by a retardation or phase delay. The ideas of conventional circuit theory, which assume that the effects of currents and charges are felt instantaneously over all the circuit, are practically exact if we confine ourselves to circuits or regions small compared with wavelength. In a large portion of the problems of modern radio engineering, the discussion cannot be restricted to such small regions. A study of the fields in the region between a transmitting antenna and a receiving antenna, to mention one example, must involve a region extremely large compared to wavelength. Efficient antenna systems themselves must be at least comparable to wavelength in size. At frequencies of the order of billions of cycles per second, almost any circuit element large enough to be of practical use must have dimensions comparable with wavelength.

Certain means were indicated in Chapter 5 for correcting the low-frequency circuit ideas for large circuits. These methods, though generally difficult to use and to visualize, enable us to retain useful circuit notions for situations in which the problem is still to find the relation between what are essentially circuit parameters, such as the ratio of input voltage applied between two points to the current flow into the system, an input impedance. But not all problems, even of those that have the current flow around a loop and applied voltage as leading characters, are best studied by continually extending circuit ideas as the only and the complete attack. In a vast number of cases attention is focused much more on the fields due to the currents and charges than on these quantities themselves. In addition we find our attention in such problems focused also on the retardation effect. Finally, it is often found necessary and desirable to use conventional circuit notions as a guide, a source of convenient notions, sometimes a stop-gap, but decidedly only as one branch of the whole of electromagnetics. Another branch is the important one of traveling electromagnetic waves.

The retardation effect leads directly to the regarding of electromag-

netic effects as a wave phenomenon. For when currents and charges change with time, the fields which they cause also change, but with a time delay that depends upon the choice of the distances between the point at which fields are being determined and the points at which the various charges and currents are located. Thus the effect of this change travels outward from the charges or currents with a finite velocity, depending upon the configuration of the conductors, and the dielectric constant and permeability of the surrounding medium. This is much the same situation as that in the transmission lines studied in Chapter 1, for a change in current or voltage at one point of a line is not felt instantaneously over all the line. Instead, it causes an effect which travels away from the point of change with a finite velocity, depending upon the distributed inductance and capacitance of the transmission line.

Waves propagate along a transmission line, according to the simple concepts described in Chapter 1, because a change in current in the line produces a voltage drop through the distributed inductance of the line, and a change in voltage produces a current through the distributed capacitance of the line. Similarly, now that displacement currents are included in the equations, it is apparent that a change in electric field produces a magnetic field in any dielectric medium; through Faraday's law we know that a change in magnetic field produces an electric field in any medium with finite permeability. By this analogy, such a wave propagation of electric and magnetic fields through any medium with finite permeability and dielectric constant should well be expected. This analogy between transmission line waves and waves in space will be seen to be a very complete one and allows us to apply directly many of the concepts of energy transmission and reflection developed for transmission lines to the study of waves in general.

The reader may well ask at this point how we can bring legitimately into a discussion of retardation electromagnetics the transmission line theory outlined in Chapter 1, when that whole study was based on circuit analysis of the conventional type. It is a good question and one that will be answered when Maxwell's equations are applied to transmission lines; but first skill must be developed in the use of wave ideas to solve electromagnetic field problems.

The retardation effect, when included in the study of circuits, was found to result in an energy term in addition to that arising from ohmic dissipation. It is a fair guess that this energy leaves the circuit in the form of waves. The matter of wave production and propagation and the behavior of these waves in and around conductors and dielectrics are both parts of a single problem. For convenience in thinking and analysis, the propagation of electromagnetic waves far from conductors will

first be studied. This will guide our approach to waves traveling along conductors such as transmission lines. Each of these will be an aid to a thorough understanding of the mechanism by which generators attached to conductors may first produce waves along conductors, and then waves in space. Thus finally, electromagnetic waves will be understood whether concentrated in a relatively closed path or region (as in circuits), flowing along conducting guides (as in transmission lines or wave guides), propagating without benefit of guiding boundaries (as in waves in free space, far from transmitter, receiver, or the earth), or transferred by conducting boundaries from a source to propagation in free space (as in an antenna).

This first chapter on wave study will be devoted to the ideas of wave propagation in unbounded media and reflection of this wave energy at discontinuities. This theory applies directly to the propagation of radio waves in space, and their reflection from dielectric, conducting, and semiconducting objects such as the earth. It will, in addition, form the foundation for later study of waves guided or enclosed by all forms of conducting and dielectric boundaries, for it will develop pictures of all boundary condition problems. The wave concepts of this and the two following chapters, quite apart from the electromagnetics, are largely applications and extensions of those built up in the first chapter on oscillations and waves. The mathematics of these chapters consists in the solution of a single differential equation, the wave equation, subject to the initial conditions describing the manner in which the wave was originated, and the boundary conditions imposed upon it by the dielectric and conducting media.

WAVES IN UNBOUNDED REGIONS

7.02 The Wave Equation Governing Electric and Magnetic Phenomena in a Charge-Free Dielectric

It has been stated that electromagnetic phenomena in free space may frequently best be regarded as wave phenomena. Now Maxwell's equations must be applied to a free dielectric to see the quantitative nature of these effects. Consider a dielectric containing no charges and with zero conductivity so that there are no conduction currents in the dielectric. The field equations are then (Art. 4.24)

$$\nabla \cdot \bar{D} = 0 \qquad [1]$$

$$\nabla \cdot \bar{B} = 0 \qquad [2]$$

$$\nabla \times \bar{E} = -\frac{\partial \bar{B}}{\partial t} \qquad [3]$$

$$\nabla \times \bar{H} = \frac{\partial \bar{D}}{\partial t} \tag{4}$$

$$\bar{B} = \mu \bar{H} \tag{5}$$

$$\bar{D} = \epsilon \bar{E} \tag{6}$$

Notice that for completeness equations (1) and (2) have been included, showing zero divergence of the fields, although they are not required if the interest is in steady state A-C components (Art. 4.23). If the dielectric is homogeneous, isotropic, and linear, ϵ and μ are constants and do not have space or time derivatives.

To attempt a solution of a group of simultaneous equations, it is usually a good plan to separate the various functions of space, such as $\bar{D}$ and $\bar{B}$, to arrive at equations that give the distributions of each.

First let us take the curl of (3)

$$\nabla \times \nabla \times \bar{E} = -\mu \nabla \times \frac{\partial \bar{H}}{\partial t}$$

Then, expanding,

$$\nabla(\nabla \cdot \bar{E}) - \nabla^2 \bar{E} = -\mu \nabla \times \left(\frac{\partial \bar{H}}{\partial t}\right)$$

By remembering that $\nabla \cdot \bar{E} = 0$ from (1) and that time and space partial derivatives may be taken in any order, and obtaining $\nabla \times \bar{H}$ from (4), we find

$$-\nabla^2 \bar{E} = -\mu \frac{\partial}{\partial t}\left(\epsilon \frac{\partial \bar{E}}{\partial t}\right)$$

or

$$\nabla^2 \bar{E} = \mu\epsilon \frac{\partial^2 \bar{E}}{\partial t^2} \tag{7}$$

This is the general form of the wave equation. The form studied in Chapter 1 was a simpler special case for one-dimensional scalars. The form of (7) applies as well to the magnetic field, as is readily shown by taking first the curl of (4) and then substituting (2) and (3).

$$\nabla^2 \bar{H} = \mu\epsilon \frac{\partial^2 \bar{H}}{\partial t^2} \tag{8}$$

From the simple special case of space variation in one dimension only, many of the characteristics of electromagnetic waves can be found that will aid in studying more complex cases. If variation is only in the z direction, the equation is simply

$$\frac{\partial^2 \bar{H}}{\partial z^2} = \mu\epsilon \frac{\partial^2 \bar{H}}{\partial t^2} \tag{9}$$

and this equation was found in Chapter 1 to have a general solution of the form

$$\bar{H} = \bar{f}_1\left(t - \frac{z}{v}\right) + \bar{f}_2\left(t + \frac{z}{v}\right) \tag{10}$$

where

$$v = \frac{1}{\sqrt{\mu\epsilon}}$$

The first term of (10) represents the wave or function $\bar{f}_1$ traveling with velocity v and unchanging form in the positive z direction; the second term represents the wave or function $\bar{f}_2$ traveling with velocity v and unchanging form in the negative z direction. It will be helpful to anticipate following discussions by pointing out that the commonest radio waves at some distance from the antenna and the ground are approximately of this simple form with space variations in one direction only.

For more general cases involving variations in more than one direction, the solution of the wave equation is not quite so simple, yet the general idea of waves propagating with definite velocities can always be obtained from it. Many of these more complicated cases will be treated later.

Problem 7.02(*a*). Show that the wave equation of the form of Eq. 7.02(7) or 7.02(8), applies to scalar potential Φ and vector potential $\bar{A}$ in a charge-free dielectric.

Problem 7.02(*b*). Show that the wave equation may be written directly in terms of any of the components of $\bar{H}$, $\bar{E}$, or $\bar{A}$ in rectangular coordinates, or to the axial component of $\bar{H}$, $\bar{E}$, or $\bar{A}$ in any coordinate system, but not to other components, such as radial and tangential components in cylindrical coordinates, or any component in spherical coordinates. That is,

$$\nabla^2 E_x = \mu\epsilon\frac{\partial^2 E_x}{\partial t^2} \qquad \nabla^2 H_z = \mu\epsilon\frac{\partial^2 H_z}{\partial t^2}, \text{ etc.}$$

but

$$\nabla^2 E_r \neq \mu\epsilon\frac{\partial^2 E_r}{\partial t^2} \qquad \nabla^2 H_\phi \neq \mu\epsilon\frac{\partial^2 H_\phi}{\partial t^2}, \text{ etc.}$$

7.03 Poynting's Theorem for Dielectric Regions

Simple transmission line waves were primarily of interest because of their ability to transfer energy from one point to another. We shall see now from the basic equations that all electromagnetic waves are capable of energy transfer. The amount and manner of this energy transfer depends somehow upon the amount, distribution, and phases of the electric and magnetic fields in the wave. This dependence will now be investigated.

Consider a region of a perfect charge-free dielectric which contains no

sources of electromagnetic energy. Since the dielectric is assumed perfect (no conductivity) there can be no dissipation of energy through current flow. Since there are no charges, no energy can be transferred to kinetic energy of motion in the charges. If there is a change in the stored energy in the electric and magnetic fields in this region, it must then have come about by an energy flow through the surface enclosing the region. The amount of energy stored in the electric fields of the region[1] is (Art. 4.24)

$$U_E = \int_V \frac{\epsilon E^2}{2}\, dV \qquad [1]$$

That stored in magnetic fields[1] is

$$U_H = \int_V \frac{\mu}{2} H^2\, dV \qquad [2]$$

The negative of the time rate of change of energy stored in the volume, which we have agreed must be the energy flowing out of the region per unit time, is

$$W = -\frac{\partial}{\partial t}(U_E + U_H) = -\frac{\partial}{\partial t}\int_V \tfrac{1}{2}(\epsilon E^2 + \mu H^2)\, dV \qquad [3]$$

The partial derivative with time may be taken inside the volume integral, but note that

$$\frac{1}{2}\frac{\partial}{\partial t}(\epsilon E^2) = E\frac{\partial(\epsilon E)}{\partial t} = \bar{E}\cdot\frac{\partial \bar{D}}{\partial t}$$

From Maxwell's equations for a charge-free dielectric

$$\frac{\partial \bar{D}}{\partial t} = \nabla \times \bar{H}$$

so

$$\frac{1}{2}\frac{\partial}{\partial t}(\epsilon E^2) = \bar{E}\cdot(\nabla \times \bar{H})$$

Similarly,

$$\frac{1}{2}\frac{\partial}{\partial t}(\mu H^2) = -\bar{H}\cdot(\nabla \times \bar{E})$$

and

$$W = \int_V [\bar{H}\cdot(\nabla \times \bar{E}) - \bar{E}\cdot(\nabla \times \bar{H})]\, dV$$

[1] As to the applicability of these expressions to time-varying phenomena, see Stratton, "Electromagnetic Theory," McGraw-Hill, 1941, pages 131–135.

The factor within the brackets is identified as $\nabla \cdot (\bar{E} \times \bar{H})$ (Art. 2.38). So

$$W = \int_V \nabla \cdot (\bar{E} \times \bar{H})\, dV$$

From the divergence theorem, the volume integral of div $(\bar{E} \times \bar{H})$ must be equal to the surface integral of $(\bar{E} \times \bar{H})$ over the surrounding boundary. Thus

$$W = \int_S (\bar{E} \times \bar{H}) \cdot \overline{dS} = \int_S \bar{P} \cdot \overline{dS} \qquad [4]$$

where

$$\bar{P} = \bar{E} \times \bar{H} \qquad [5]$$

The vector $\bar{P}$, defined as $\bar{E} \times \bar{H}$, is known as the Poynting vector. According to (4), the total energy flow out of the region per unit time is the surface integral of this vector over all the boundary surrounding the region. Since this has the dimensions of an energy flow per unit time, it is a power flow, and in the practical units used here has the dimensions of watts.

Since total power flow is given by the surface integral of $\bar{P}$, $\bar{P}$ itself may be thought of as a surface density of power flow (watts per square meter), giving the direction and magnitude of power flow out of any volume per unit area at each point of the enclosing surface. This is a convenient concept, but it must be emphasized that it has not actually been proved by the foregoing derivation. Equation (4) proves only that if the surface integral of $\bar{P}$ is taken *over all the surrounding boundary*, the total power calculated must be that flowing out of the region; it does not necessarily follow that $\bar{P}$ is the actual density of power flow at each point. Nevertheless, we shall often use this as a concept, and so long as its limitations are realized, it will be a very useful one.

Equation (4) is a special case of the theorem known as Poynting's theorem. Equation (4) was proved for charge-free, perfect dielectric regions with no sources. Actually it is much more general than this. When we consider a region that includes the source of the waves, imperfect conductors, charges, etc., it will be shown that the net power flow out of any such region is still given by the surface integral of the Poynting vector $\bar{E} \times \bar{H}$ over the surrounding boundary.

7.04 Uniform Plane Waves in a Perfect Dielectric

Consider now the simple case in which there are no variations except in one direction so that the wave equation has the form written in

Eq. 7.02(9) and has solutions of the form of Eq. 7.02(10). We have already identified the two possible solutions as waves propagating in the positive z direction and negative z direction respectively. If, due to some originating cause which is not under discussion here, a wave has been started in the positive z direction in an unbounded medium, and there is nowhere any discontinuity or object that might cause a reflected wave, the wave traveling in the negative z direction will not appear in the solution. There is only

$$\bar{H} = \bar{f}\left(t - \frac{z}{v}\right) \qquad [1]$$

If the source in the plane $z = 0$ is one that causes variations sinusoidal in time, then at $z = 0$

$$\bar{H}\Big|_{z=0} = \bar{H}_0\, e^{j\omega t}$$

At any plane z, from (1),

$$\bar{H} = \bar{H}_0\, e^{j\omega\left(t - \frac{z}{v}\right)} \qquad [2]$$

For the study of sinusoidal time variations, it will be convenient to have the factor $e^{j\omega t}$ understood, so it will henceforth not be written explicitly. Moreover, certain quantities, propagation constant γ and phase constant β, will prove useful which for the waves under discussion now are given by

$$\gamma = j\beta = \frac{j\omega}{v} = j\omega\sqrt{\mu\epsilon} \qquad [3]$$

It will often be convenient to allow the factor $e^{-j\beta z}$ to be understood as well. Thus in speaking of waves with a propagation constant γ, it will be understood that all quantities are multiplied by the propagation function $e^{(j\omega t - \gamma z)}$. Just as time derivatives were replaced previously by $j\omega$, z derivatives may then be replaced by $-\gamma$.

With the above conventions and by remembering that the present discussion concerns the type of wave which is uniform over the x-y plane (so that $\partial/\partial x$ and $\partial/\partial y$ are equal to zero), Eq. 7.02(3) divides into the following component equations:

$$\gamma E_y = -j\omega\mu H_x \quad \text{or} \quad E_y = \frac{-j\omega\mu H_x}{j\omega\sqrt{\mu\epsilon}} = -\eta H_x \qquad [4]$$

$$-\gamma E_x = -j\omega\mu H_y \quad \text{or} \quad E_x = \eta H_y \qquad [5]$$

$$0 = -j\omega\mu H_z \qquad [6]$$

where

$$\eta = \sqrt{\frac{\mu}{\epsilon}}$$

Equation 7.02(4) gives the same relations as (4) and (5) and the additional information that

$$j\omega\epsilon E_z = 0 \qquad [7]$$

The above equations show first, (6) and (7), that there is no component of either electric field or magnetic field in the direction of propagation for such a uniform plane wave. A study of (4) and (5) shows that total electric field and total magnetic field ($\bar{E}$ and $\bar{H}$) are in time phase, mutually perpendicular at every point of the wave and related in magnitude by the quantity η. This quantity has the dimensions of an impedance (ohms) and is called the *intrinsic impedance* of the medium. Had a negatively directed wave been considered instead of the positively traveling one, the equations (4) to (7) would alter only in certain signs; that is, γ would now become negative and would initiate these sign changes. $\bar{E} \times \bar{H}$ is entirely in the direction of propagation. Had there been a component of electric field or of magnetic field in the direction of propagation, $\bar{E} \times \bar{H}$ would have had a component normal to the direction of propagation. According to the Poynting theorem this would have represented energy transfer in a direction other than the direction of propagation.

The stored energy per unit volume of the dielectric is, in magnetic energy,

$$U_H = \frac{\mu}{2} H^2 = \frac{\mu}{2} (H_x^2 + H_y^2)$$

and in electric energy,

$$U_E = \frac{\epsilon E^2}{2} = \frac{\epsilon}{2} (E_x^2 + E_y^2) = \frac{\epsilon\mu}{2\epsilon} (H_y^2 + H_x^2)$$

The energy stored in electric field per unit volume at any point of the wave is then equal at every instant to the energy stored in magnetic field per unit volume at that point. The exact behavior of energy stored in a given volume of space can be visualized best by a study of the energy flow in and out of that volume as a result of the wave action. If the Poynting theorem, Eq. 7.03(4), is applied to a small rectangular parallelepiped aligned with the direction of propagation (Fig. 7.04), there are no contributions to $\bar{P}$ through any of the sides except the ends $ABCD$

and $EFGH$, since $\bar{P}$ is exactly in the direction of propagation. The flow through each of these sides at any instant is merely

$$W = \text{Area} \times (\bar{E}_i \times \bar{H}_i)$$

where $\bar{E}_i$ amd $\bar{H}_i$ are instantaneous values of $\bar{E}$ and $\bar{H}$. At any instant, the power flow through one of these ends will not be the same as that out of the other unless the ends are separated by a multiple of a half wavelength ($\beta z_0 = n\pi$). In particular, if the ends are separated by an odd multiple of a quarter wavelength $\left(\beta z_0 = \dfrac{(2n+1)\pi}{4}\right)$, the power flow through $ABCD$ is a maximum at the instant the flow through $EFGH$ is zero, and there is a net increase of stored energy in the parallelepiped. A quarter of a cycle later in time the flow through $EFGH$ is a maximum, that through $ABCD$ is zero, and there is a net decrease in the energy stored in the parallelepiped. Note though that there is never a reversal of the Poynting vector $\bar{P}$; it always points in the positive z direction since $\bar{E}$ and $\bar{H}$ are in phase, and both change signs at the same instant. Note also that the *time average value* of power flow through any surfaces, as $ABCD$ and $EFGH$, must be the same no matter what the spacing z_0. With wave amplitudes E_x, E_y, H_x, and H_y, this time average power flow per unit area is

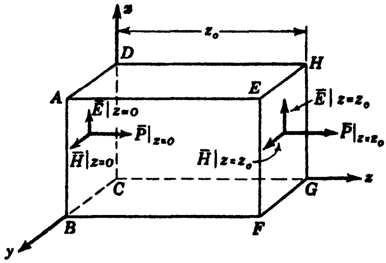


FIG. 7.04. Poynting flow through a rectangular parallelepiped aligned with direction of propagation.

$$P_z = \text{Time average } (\bar{E} \times \bar{H})_z = \tfrac{1}{2}(E_xH_y - E_yH_x)$$

$$= \frac{\eta}{2}(H_y^2 + H_x^2) = \frac{\eta H^2}{2} = \frac{E^2}{2\eta} \quad \text{watts/meter}^2$$

For this simple wave, all points in a plane normal to direction of propagation are in time phase. This is the reason for calling such a wave a plane wave. We shall later study waves which are not of this type, such as spherical waves and cylindrical waves. Since there are also no variations in magnitude in the plane normal to direction of propagation for this simple wave, it may then be further called a *uniform plane wave.* In studies of the wave propagation along transmission lines and dielectric or conducting wave guides we shall find plane waves which are not uniform. This uniform plane wave may also be called a *transverse electromagnetic wave* if desired, since it has no electric and magnetic field

components except those in the plane normal (that is, transverse) to the direction of propagation. There are also other transverse electromagnetic waves which are not uniform plane waves.

Below are summarized the properties found for the uniform plane wave.

1. Velocity of propagation, $v = 1/\sqrt{\mu\epsilon}$.
2. No electric or magnetic field in direction of propagation.
3. Total electric field normal to total magnetic field.
4. Electric field in time phase with magnetic field.
5. Magnitudes of electric and magnetic field related by

$$E = \eta H$$

6. Direction of propagation given by direction of Poynting vector.

$$\bar{P} = \bar{E} \times \bar{H}$$

7. Energy stored in electric field per unit volume at any instant and any point is equal to energy stored in magnetic fields per unit volume at that instant and that point.

8. Average power flow per unit area through a plane perpendicular to the direction of propagation is

$$P_z = \frac{\eta}{2} H^2 = \frac{E^2}{2\eta} \quad \text{watts/meter}^2$$

where the maximum values of the instantaneous fields at any point are E and H.

7.05 Combinations of Uniform Plane Waves — Polarization

Since the wave equation is a linear equation, any solution to it may be built up as the sum of other solutions. Many complex electromagnetic wave distributions might, if desired, be considered as made up of a large number of the simple plane waves with different magnitudes, phases, and directions of propagation. For most purposes this viewpoint is of little value except as a concept, and other methods to be given later will serve better for actual analysis. However, if we are studying the important practical case where a combination of plane waves exists such that all have the same direction of propagation, there is a definite advantage in considering these as a superposition of the individual plane waves and analyzing by obtaining the behavior of each individual wave. The orientations of the field vectors in these waves are often described by the *polarization* of the wave.

For a single uniform plane wave, it has been seen that electric and magnetic field vectors are always at right angles and always maintain

their respective orientations at every point along the wave. A combination of plane waves all propagating in the same direction, and with arbitrary orientations of the field vectors, is called an *unpolarized* wave. These individual component waves are often of different magnitudes and phases as well (Fig. 7.05*a*).

If all the plane waves propagating in the same direction have the same orientation of the field vectors, the wave is said to be *plane polarized.* A question may arise here as to whether these definitions have to do with only one frequency or not. In radio engineering, where the commonest dielectric is free space, the propagation constant is independent of fre- ?

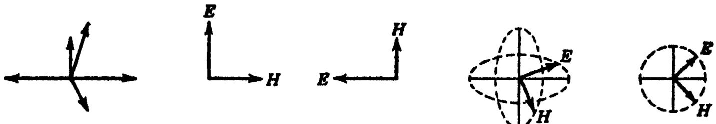


FIG 7 05

quency and the definitions so far stated will be unaffected by frequency considerations. However, for a wave made up of two different frequencies with the same propagation direction and same orientation of field vectors, one would probably not speak of a plane polarized wave but rather two plane polarized waves of same polarization, different frequencies.

In radio engineering, the plane of polarization is customarily defined by the orientation of the electric field vector, although in optics the convention is that the magnetic field vector defines the plane of polarization. Thus a combination of waves all with electric field vector in the vertical plane is said to be *vertically polarized* according to radio engineering terminology. If electric vectors of all waves are in the horizontal plane, there is said to be *horizontal polarization* (Fig. 7.05*b* and *c*).

If there is a combination of two uniform plane waves of the same frequency, but of different phases, magnitudes, and orientations of the field vectors, the resultant combination is said to be an *elliptically polarized* wave. To see the reason for this, we may first break each wave into its two separate component waves, one with electric vector in the x direction, the other with electric vector in the y direction. The two x components add to produce a wave of given magnitude and phase angle. This will be written for this study directly in terms of cosines rather than the exponential or complex form.

$$E_x = E_1 \cos \omega \left(t - \frac{z}{v}\right)$$

The two y components add to produce a wave of different magnitude and phase angle.

$$E_y = E_2 \cos\left[\omega\left(t - \frac{z}{v}\right) + \psi\right]$$

In any given plane, say $z = 0$, these reduce to equations of the form

$$E_x = E_1 \cos \omega t$$

$$E_y = E_2 \cos (\omega t + \psi)$$

These are the parametric equations for an ellipse. The terminus of the electric field vector then traces an elliptic path in a plane normal to the direction of propagation. This is the reason for the name elliptic polarization (Fig. 7.05*d*).

If the two waves above combine so that total x and y components are equal and 90° out of time phase, the ellipse reduces to a circle, and the wave is said to be *circularly polarized.* Thus if

$$\psi = \frac{\pi}{2} \quad \text{and} \quad E_1 = E_2$$

$$E_x^2 + E_y^2 = E_1^2$$

which is the equation of a circle (Fig. 7.05*e*).

REFLECTION OF WAVES FROM CONDUCTORS AND DIELECTRICS; THE IMPEDANCE CONCEPT

7.06 Reflection of Normally Incident Plane Waves from Perfect Conductors

If a uniform plane wave, as studied in Art. 7.04, is imagined as a process of energy flow through space, then, in considering a plane perfect conductor lying normal to the direction of propagation, we feel instinctively that there will be a steady stream of reflected waves resulting from the incidence of the initial waves on this plate. It will no longer be possible to describe the fields in front of the conductor by the single function of $t - (z/v)$. There should now, it seems, be present also a wave which travels in the negative z direction; that is, a function of $t + (z/v)$.

For one thing, we cannot satisfy the required boundary conditions of zero electric field at a given value of z for all values of time with the single wave function. Since the fields in free space can be expressed as a sum of the two functions, it is seen immediately that to satisfy the

boundary conditions there will be just enough of each function so that the resultant tangential electric field at the conductor's surface will be zero.

From another viewpoint, it is readily realized from the Poynting theorem that no energy can pass the surface of a perfect conductor. This is true because a perfect conductor requires that the component of $\bar{E}$ tangential to the conducting surface must be zero, and thus there can be no component of Poynting vector, $\bar{P} = \bar{E} \times \bar{H}$, normal to the perfect conducting surface. All energy associated with the incident wave must then be reflected in some manner. Thus it appears that there is in addition to the incident wave, a reflected wave traveling in the negative z direction, equal in magnitude to the incident wave if it is to contain all the energy brought by the incident wave.

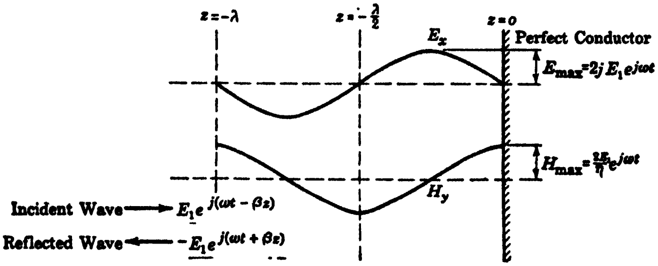


FIG. 7.06. Reflection of a uniform plane wave from a perfect conductor.

For a single plane wave, select the orientation of axes so that total electric field lies in the x direction and include waves traveling both in the positive and negative z directions (Fig. 7.06).

$$E_x = Ee^{j(\omega t-\beta z)} + E'e^{j(\omega t+\beta z)}$$

If $E_x = 0$ at $z = 0$ for all values of time, $E' = -E$.

$$E_x = E(e^{-j\beta z} - e^{j\beta z})e^{j\omega t}$$

but

$$e^{jx} = \cos x + j \sin x$$

So

$$E_x = -2jE \sin \beta z \; e^{j\omega t} \qquad [1]$$

From Eq. 7.04(5) $H_y = \dfrac{E_x}{\eta}$ for a wave traveling in the positive z direction. For a wave traveling in the negative z direction Eq. 7.04(5)

would show that $H_y = -E_x/\eta$.

$$H_y = \left(\frac{E}{\eta} e^{j(\omega t - \beta z)} - \frac{E'}{\eta} e^{j(\omega t + \beta z)}\right)$$

$$= \frac{E}{\eta} (e^{-j\beta z} + e^{j\beta z}) e^{j\omega t}$$

$$= \frac{2E}{\eta} \cos \beta z \, e^{j\omega t} \qquad [2]$$

Equations (1) and (2) state that although total electric and magnetic fields for the combination of incident and reflected waves are still mutually perpendicular in space and related in magnitude by η, they are now in time quadrature. The pattern is a standing wave pattern since a zero of electric field is always at the conductor surface, and also always at $\beta z = -n\pi$ or $z = -n\lambda/2$. Magnetic field has a maximum at the conductor surface, and there are other maxima each time there are zeros of electric field. Similarly, zeros of magnetic field and maxima of electric field are at $\beta z = \dfrac{2n+1}{2}\pi$, or $z = -\dfrac{(2n+1)\lambda}{4}$. This situation is sketched in Fig. 7.06, a typical standing wave pattern such as was found for the shorted transmission line in Chapter 1. At an instant in time, occurring twice each cycle, all the energy of the line is in the magnetic field; 90° later the energy is stored entirely in the electric field.

7.07 Transmission Line Analogy of Wave Propagation; the Impedance Concept

The resemblance between the standing wave patterns obtained when an electromagnetic wave impinges upon a perfect conductor and for the wave in a shorted transmission line is but one indication of the complete analogy between the two phenomena. This article will show the basic character of this analogy and some of the uses to which it may be put. This will only be, however, the beginning of the discussion of the analogy between electromagnetic wave behavior and the behavior of circuits and lines as analyzed by familiar circuit concepts such as *impedance*. Such discussion will be continued throughout the remainder of the text until, as electromagnetic wave phenomena become understood in all possible roles, the relation of wave and field ideas to the conventional circuit and line ideas is put on a clear and rigorous basis.

For the present, the points of similarity between the results of the preceding rigorous analyses of plane waves and the conventional analysis of transmission lines will be pointed out.

1. The velocity of propagation of a wave in a uniform transmission line was found in terms of the distributed constants of the line, inductance and capacitance per unit length:

$$v = \frac{1}{\sqrt{LC}} \quad \text{meters/sec} \tag{1}$$

The velocity of propagation of a uniform plane wave was found in terms of permeability and dielectric constant:

$$v = \frac{1}{\sqrt{\mu\epsilon}} \tag{2}$$

where μ has units of distributed inductance, henrys/meter, and ϵ has units of distributed capacitance, farads/meter.

2. In a transmission line, voltage and current associated with a single wave at every point are related by the characteristic impedance of the line:

$$\frac{V}{I} = Z_0 = \sqrt{\frac{L}{C}} \quad \text{ohms} \tag{3}$$

For a wave in the negative direction:

$$\frac{V}{I} = -\sqrt{\frac{L}{C}}$$

For a uniform plane wave in an unbounded medium, electric field and magnetic field associated with a single wave are related at every point by the intrinsic impedance of the medium:

$$\frac{E}{H} = \frac{E_x}{H_y} = -\frac{E_y}{H_x} = \eta; \quad \eta = \sqrt{\frac{\mu}{\epsilon}} \quad \text{ohms} \tag{4}$$

For a wave in the negative direction:

$$\frac{E_x}{H_y} = -\frac{E_y}{H_x} = -\eta$$

3. For an impedance change at a point in the transmission line, the amounts of transmitted and reflected waves were determined from the conditions that current and voltage must be continuous (equal) on the two sides of the discontinuity.

For an electromagnetic wave passing from one medium to another, the amounts of transmitted and reflected waves are determined from the conditions that tangential electric and magnetic fields are continuous on the two sides of the discontinuity.

To complete the analogy rigorously, consider the plane boundary between any two media (Fig. 7.07*a*). Suppose there is a wave in medium 1, $E_1 e^{(j\omega t - \gamma_1 z)}$ incident upon the boundary. There is a reflected wave $E_1' e^{(j\omega t + \gamma_1 z)}$, and a wave transmitted into medium 2, $E_2 e^{(j\omega t - \gamma_2 z)}$. Take the boundary as $z = 0$; total tangential components of electric and magnetic field should be equal on the two sides of this boundary:

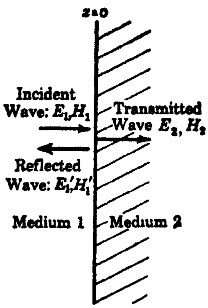


FIG. 7.07*a*. Reflection and transmission at a plane boundary between two media.

$$E_{t1} + E_{t1}' = E_{t2} \tag{5}$$

$$H_{t1} + H_{t1}' = H_{t2} \tag{6}$$

If in medium 1 E_{t1}/H_{t1} = constant = $Z_1 = -E_{t1}'/H_{t1}'$ and in medium 2, $E_{t2}/H_{t2} = Z_2$, (5) and (6) give

$$\frac{E_{t1}'}{E_{t1}} = -\frac{H_{t1}'}{H_{t1}} = \frac{Z_2 - Z_1}{Z_2 + Z_1} = \frac{K-1}{K+1} \tag{7}$$

and

$$\frac{E_{t2}}{E_{t1}} = K\frac{H_{t2}}{H_{t1}} = \frac{2Z_2}{Z_2 + Z_1} = \frac{2K}{K+1} \tag{8}$$

where

$$K = \frac{Z_2}{Z_1} \tag{9}$$

These are completely similar to the equations of Art. 1.18, giving transmitted and reflected current and voltage waves in a transmission line.

The condition of constant ratio of transverse components of electric and magnetic fields over the boundary is apparently automatically satisfied for a uniform plane wave normally incident upon the boundary; the constant is η. However, in addition to these simple uniform plane waves there are other waves which also will be found to fulfil the requirement of constant ratio between tangential electric and magnetic field components, and for these the impedance concepts and transmission line analogies will also be found valid and useful. In general, for all such waves, the wave impedance will always mean the ratio of tangential electric field to tangential magnetic field. The convention as to signs can be chosen arbitrarily to suit convenience as long as consistency is maintained. The sign in the impedance relation will be taken positive if the direction of $\bar{E}$, $\bar{H}$, and the positive direction of propagation follow each other in the order of the coordinates. Thus if the tangential electric field is in the x direction, tangential magnetic field in the y direction,

the wave impedance for a wave propagating in the positive z direction is

$$\frac{E_x}{H_y} = Z_z$$

If a wave, on the other hand, has as tangential fields E_y and H_x, the wave impedance looking in the positive z direction will be expressed by

$$\frac{E_y}{H_x} = -Z_z$$

Similarly, if the subscript on Z is always understood to mean that the direction of that coordinate is considered as the positive direction of propagation,

$$\frac{E_y}{H_z} = Z_x \qquad \frac{E_z}{H_x} = Z_y$$

$$\frac{E_z}{H_y} = -Z_x \qquad \frac{E_x}{H_z} = -Z_y$$

For impedances looking in the negative x, y, or z directions, the signs in these equations will of course be reversed.

In a region that contains both incident and reflected waves, the total impedance at any plane (ratio of total transverse components of E and H) may be obtained in terms of the length of the region, propagation constant through the region, and the terminating impedance of the region (Fig. 7.07b). The result is exactly similar to that developed for input impedance of a transmission line of general length, propagation constant, and terminating impedance.

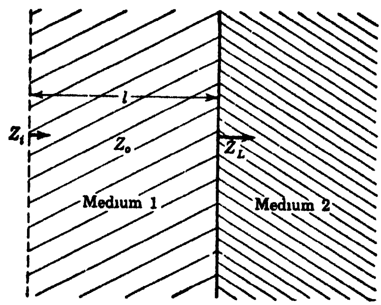


FIG. 7.07b. Region of wave propagation terminated by a second medium.

$$Z_i = Z_0 \left[\frac{Z_L \cos \beta l + jZ_0 \sin \beta l}{Z_0 \cos \beta l + jZ_L \sin \beta l} \right] \qquad [10]$$

where l = length of line or region of wave propagation.
Z_i = input impedance.
Z_L = terminating or load impedance.
β = phase constant.
Z_0 = characteristic impedance, or ratio of transverse electric field to transverse magnetic field for a single wave.

It is quite evident that the impedance concept does not have to be applied to electromagnetic waves and that answers to wave problems may be obtained without pointing out the analogy to conventional transmission line analysis. The situation is similar to that of the concept of $R + jX$ in A-C circuit analysis which is never *necessary* to the solution of an A-C problem but which is of undoubted value in thinking and computing. By making proper use of the impedance concept, not only can results from one field be used directly in the other but also the *relation* of one branch of electromagnetics to the other becomes clearer. Credit for properly evaluating the importance of the wave impedance concept to engineers and making its use clear belongs to S. A. Schelkunoff.[2]

7.08 Normal Incidence on Perfect Dielectric

If a uniform plane wave is normally incident upon a perfect dielectric, the transmission line analogy of Art. 7.07 may be applied at once. This dielectric is assumed to be infinite in extent beyond the boundary so that no multiple reflections need be present for this discussion. Select the direction of electric field as the direction of the x axis, and direction of propagation for the incident wave as the positive z direction. Then in the incident wave there are present only E_x and H_y, and from Eq. 7.04(5)

$$\frac{E_{x1}}{H_{y1}} = \eta_1 = \sqrt{\frac{\mu_1}{\epsilon_1}} \qquad [1]$$

In the reflected wave

$$\frac{E'_{x1}}{H'_{y1}} = -\eta_1 \qquad [2]$$

In the wave transmitted to the second dielectric,

$$\frac{E_{x2}}{H_{y2}} = \eta_2 = \sqrt{\frac{\mu_2}{\epsilon_2}} \qquad [3]$$

Then by Eq. 7.07(7) and (8), if $K = \eta_2/\eta_1$

$$\frac{E'_{x1}}{E_{x1}} = -\frac{H'_{y1}}{H_{y1}} = \frac{K-1}{K+1} \qquad [4]$$

$$\frac{E_{x2}}{E_{x1}} = K\frac{H_{y2}}{H_{y1}} = \frac{2K}{K+1} \qquad [5]$$

Note that there is perfect transmission and no reflected wave at all if $K = 1$, that is, if $\eta_1 = \eta_2$. If $K \ll 1\,(\eta_1 \gg \eta_2)$ there is very little energy

[2] See, for instance, *Bell System Tech. Journ.*, 17, 17, 1938.

transmitted; almost all is reflected. Moreover, the phase relations of the reflected wave are exactly the same as those found in Art. 7.06 for reflection from a perfect conductor; the standing wave patterns are consequently the same as in that case, with a node of electric field and a maximum of magnetic field at the reflecting surface. This is true of a wave passing from one dielectric to a second dielectric of much greater dielectric constant or smaller permeability. The very high displacement currents at the surface of this second dielectric then have the same effect in shorting the electric field and hence producing the wave pattern that includes a node in $\bar{E}$, as do the conduction currents in the perfect conductor.

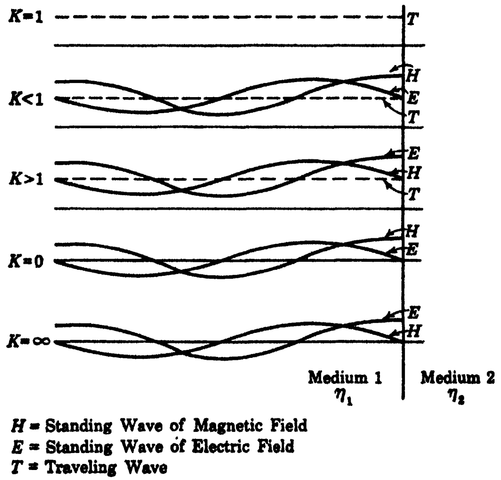


FIG. 7.08. Standing and traveling waves upon change of medium ($K = \eta_2/\eta_1$).

If $K \gg 1 (\eta_2 \gg \eta_1)$, reflection is again almost perfect, but now the phase of the reflected wave is opposite to that of the previous case, and there is a maximum of electric field and a minimum of magnetic field at the surface. This corresponds to a wave passing from one dielectric to another of much smaller dielectric constant or greater permeability.

In any intermediate case for K neither unity, nor approaching zero or infinity, the wave is partially reflected and partially transmitted. The result is then a combination of a traveling wave plus a standing wave in the region of the first medium. The traveling wave portion corresponds to the amount of energy transmitted to the second dielectric, and the standing wave corresponds to the portion of the incident wave that is reflected. This latter standing wave has zeros and maxima exactly at half-wave points as before, but there are now no points in region 1

where the *total* field is always zero since the traveling wave must be added. For values of K less than 1, the standing wave component has a zero of electric field and a maximum of magnetic field at the discontinuity; for values of K greater than 1, the reverse is true. As K approaches unity, the standing wave component becomes smaller, and of course disappears at $K = 1$.

Several of the above situations are sketched in Fig. 7.08. The complete functions describing the combinations of standing plus traveling wave in medium 1 may be written so as to emphasize the factors of transmission and reflection:

$$\frac{E_x}{E_{x1}} = 2\left[\frac{1}{K+1}e^{-j\beta z} + \frac{K-1}{K+1}\cos\beta z\right]e^{j\omega t} \quad (K > 1) \qquad [6]$$

$$\frac{H_y}{H_{y1}} = 2\left[\frac{1}{K+1}e^{-j\beta z} + j\frac{1-K}{1+K}\sin\beta z\right]e^{j\omega t} \quad (K > 1) \qquad [7]$$

$$\frac{E_x}{E_{x1}} = 2\left[\frac{K}{K+1}e^{-j\beta z} - j\frac{1-K}{1+K}\sin\beta z\right]e^{j\omega t} \quad (K < 1) \qquad [8]$$

$$\frac{H_y}{H_{y1}} = 2\left[\frac{K}{K+1}e^{-j\beta z} + \frac{1-K}{1+K}\cos\beta z\right]e^{j\omega t} \quad (K < 1) \qquad [9]$$

E_{x1} and H_{y1} represent field components in the incident wave.

7.09 Elimination of Wave Reflections for Normal Incidence on Perfect Dielectrics

It was found in the previous article that all energy is transmitted, none is reflected, if the two dielectric materials have equal intrinsic impedances, $\eta_1 = \eta_2$, or

$$\sqrt{\frac{\mu_1}{\epsilon_1}} = \sqrt{\frac{\mu_2}{\epsilon_2}} \qquad [1]$$

However, knowledge of this is of little aid in attaining perfect matching since no dielectric materials with permeabilities very much different from that of air are known, and if $\mu_1 = \mu_2$ then (1) requires that $\epsilon_1 = \epsilon_2$. This is a trivial case since it is obvious that there would be no reflection if the two materials have identical dielectric constants and permeabilities.

Recall, though (Prob. 1.21*b*), that it is possible for a single frequency to match perfectly two transmission lines of different characteristic impedances by introducing between them a quarter-wave matching section. This matching section, or impedance transformer, was a transmission line a quarter wave in length and having a characteristic imped-

ance equal to the geometric mean of the two characteristic impedances to be matched. With this matching section, the wave reflections from the two discontinuities arrive with proper magnitudes and phases to cancel completely the backward traveling wave in the original line and allow all energy to be transmitted.

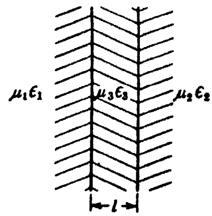


FIG. 7.09. Matching section for plane waves.

Similarly, the transmission line analogy shows that wave energy of a single frequency may be perfectly transmitted between one dielectric material and another, if a third dielectric section is introduced at the boundary (Fig. 7.09). This impedance matching region must be a quarter wave in thickness, and of intrinsic impedance the geometric mean of the two intrinsic impedances to be matched. That is,

$$l = \frac{\lambda_3}{4} = \frac{1}{4f\sqrt{\mu_3\epsilon_3}} \qquad [2]$$

and

$$\frac{\mu_3}{\epsilon_3} = \sqrt{\frac{\mu_1\mu_2}{\epsilon_1\epsilon_2}} \qquad [3]$$

7.10 Phase Velocities for Waves at Any Angle of Incidence

We have considered, in the previous articles, waves normally incident upon conductors and dielectrics. When the incidence is at any general angle, the problem may of course be solved by the usual method of matching tangential field components at the boundary. However, the previous analogies and equations may be applied directly if the viewpoint of wave propagation is broadened somewhat. In this extension, any uniform plane wave propagating with velocity of light in any given direction may be regarded as a somewhat different wave, propagating in some more favorable direction (say normal to a discontinuity) by simply altering its phase velocity in the proper manner.

Consider, in Fig. 7.10, a uniform plane wave propagating in the z' direction AO. From the previous analysis of uniform plane waves, it is known that this wave will have field components only in transverse planes, as aa or $a'a'$, and these components do not vary in magnitude or in phase along these planes. Propagation in the z' direction is with the speed of light in the dielectric (1), so that the propagation function is

$$E_1 = Ee^{j\omega\left(t - \frac{z'}{v_1}\right)} = Ee^{j(\omega t - \beta_1 z')}$$

And v_1 (or ω/β_1) is, of course, the phase velocity in the z' direction. That is, if the phase angle is to appear constant to a fictitious moving observer, his velocity would be

$$\frac{dz'}{dt} = \frac{\omega}{\beta_1}$$

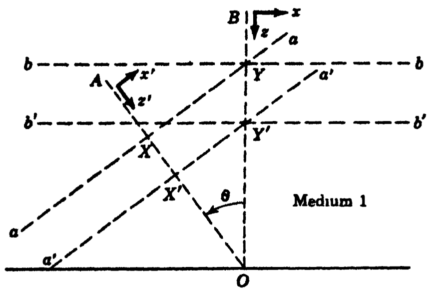


FIG. 7.10. Uniform plane wave moving at angle θ toward a plane.

Now, the fictitious point Y, representing intersections of the plane of constant phase aa and the axis BO, moves in the z direction (along BO) with velocity $v_1/\cos\theta$. To see just what is meant by this statement, consider a plane normal to the z direction. Instead of uniformity at every instant there are variations of electric and magnetic field along this plane, producing a given pattern at any instant. In the time that a plane of constant phase aa has moved to $a'a'$ over the distance XX', namely,

$$t_{xx'} = \frac{XX'}{v_1}$$

the pattern in the plane bb has moved to a new plane $b'b'$. Thus the phase velocity in the z direction v_{1z} is determined by

$$t_{yy'} = t_{xx'} = \frac{YY'}{v_{1z}}$$

But

$$XX' = YY' \cos\theta$$

So

$$v_{1z} = \frac{YY'v_1}{XX'} = \frac{v_1}{\cos\theta} \qquad [1]$$

And the propagation function in the z direction may be written

$$E_1 = E'e^{j\omega\left(t - \frac{z\cos\theta}{v_1}\right)} = E'e^{j\omega\left(t - \frac{z}{v_{1z}}\right)} \qquad [2]$$

In other words, we say that one never knows or cares whether there is a uniform plane wave traveling at angle θ toward the boundary with velocity v_1, or a non-uniform wave traveling with angle zero (normal incidence) toward that boundary with phase velocity $v_1/\cos\theta$. The non-uniformity shows up in E' which is a function (no longer a constant) in the plane normal to the newly selected direction of propagation BO.

We may obtain this function E' and, indeed, derive (1) and (2) in another way by noting in Fig. 7.10 that a distance z', in terms of the coordinates x and z, is

$$z' = z \cos \theta + x \sin \theta$$

So that the original wave

$$Ee^{j(\omega t - \beta_1 z')} = Ee^{-j\beta_1 x \sin \theta} e^{j(\omega t - \beta_1 z \cos \theta)}$$

The distribution function in the x direction, E' of (2), is then given by the $Ee^{-j\beta_1 x \sin \theta}$ term, and the phase velocity in the z direction, just as determined by the previous reasoning is merely

$$\frac{\omega}{\beta_1 \cos \theta} = \frac{v_1}{\cos \theta}$$

Note. It is equally proper to write the propagation function as

$$Ee^{-j\beta_1 z \cos \theta} e^{j(\omega t - \beta_1 x \sin \theta)}$$

and to consider the wave as a pattern $Ee^{-j\beta_1 z \cos \theta}$ propagating in the x direction with velocity

$$\frac{\omega}{\beta_1 \sin \theta} = \frac{v_1}{\sin \theta}$$

In either case, the wave pattern in the plane transverse to the selected direction of propagation is not a uniform pattern, and its points are not in phase, but so long as the *ratio* of transverse components of electric to magnetic field is a constant, the general expressions for wave reflections developed in Art. 7.07 are still valid for analysis, and the transmission line analogies are equally valid for quantitative thinking. This use will be demonstrated by examples to follow.

Problem 7.10. Find the *group* velocity of waves such as those discussed in the preceding article (uniform plane waves incident at a boundary at an arbitrary angle) in the direction of the normal to the boundary.

7.11 Incidence at Any Angle on Perfect Conductors

We shall utilize the concepts of phase velocity and non-uniform plane waves to consider a very simple case, that of incidence upon a perfect conductor at any oblique angle. No matter what the orientation of the field vectors, the wave may be broken into two components, one with the electric vector entirely in the plane of incidence, the other with the magnetic vector entirely in the plane of incidence. The first of these will be said to be polarized in the plane of incidence, the second polarized normal to the plane of incidence. These may then be considered separately.

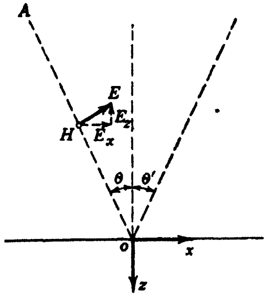


FIG. 7.11. Wave polarized with E in plane of incidence striking a plane boundary at angle θ.

A. Polarization in Plane of Incidence. In Fig. 7.11 the angle θ is measured from the normal to the surface, and the plane of incidence is chosen as the xz plane. The magnetic vector lies entirely in the y direction. Suppose its magnitude is H. According to concepts of the previous article, we can consider this incident uniform plane wave traveling in the AO direction as a non-uniform wave traveling in the z direction as follows.

$$H_{y1} = He^{-j\beta_1 x \sin\theta} e^{j(\omega t - \beta_1 z \cos\theta)} \quad [1]$$

$$E_{x1} = \eta_1 H \cos\theta \, e^{-j\beta_1 x \sin\theta} e^{j(\omega t - \beta_1 z \cos\theta)} \quad [2]$$

$$E_{z1} = -\eta_1 H \sin\theta \, e^{-j\beta_1 x \sin\theta} e^{j(\omega t - \beta_1 z \cos\theta)} \quad [3]$$

The ratio between transverse components of electric and magnetic field is a constant for all values of x, y, and z for this wave,

$$\frac{E_{x1}}{H_{y1}} = \eta_1 \cos\theta = Z_1 \quad [4]$$

so, according to previous arguments, it is quite proper to use the impedance concept with Z_1 as the characteristic wave impedance for the wave.

Since energy cannot be transferred to the perfect conductor, there must again be a reflected wave. This is reflected at some yet unknown angle θ'. To find its phase velocity along any particular direction let us look for a moment at both incident and reflected waves as propagating in the x direction. Then we see that both waves must travel at the same phase velocity in this direction because they must combine everywhere along the plane in precisely the same way to satisfy the unchanging boundary condition along that plane.

$$\frac{v_1}{\sin\theta} = \frac{v_1}{\sin\theta'}$$

Another way to state the same reasoning is that a boundary condition of zero tangential electric field is imposed by the conductor for all values of x and time. This can only be satisfied if

$$e^{-j\beta_1 x \sin\theta} = e^{-j\beta_1 x \sin\theta'}$$

So from either viewpoint it follows that the reflected ray's electric field lies in the plane of incidence; that is, there can be no y component of propagation. It is also necessary that the angle of reflection be equal to the angle of incidence.

$$\theta' = \theta \quad [5]$$

So for the reflected wave

$$E'_{x1} = -\eta_1 H'_y \cos\theta\, e^{-j\beta_1 x \sin\theta}\, e^{j(\omega t + \beta_1 z \cos\theta)} \qquad [6]$$

$$E'_{z1} = -\eta_1 H'_y \sin\theta\, e^{-j\beta_1 x \sin\theta}\, e^{j(\omega t + \beta_1 z \cos\theta)} \qquad [7]$$

The wave impedance, or ratio of transverse electric to magnetic field in this wave, is the negative of that for the incident wave because of the difference in direction.

$$Z'_1 = \frac{E'_{x1}}{H'_y} = -\eta_1 \cos\theta = -Z_1$$

This is another of the requirements for the results of Art. 7.07 to apply here. The wave impedance of the reflecting medium must be zero since the perfect conductor requires zero tangential electric field. So from Eqs. 7.07(7) and 7.07(9)

$$K = \frac{Z_2}{Z_1} = 0$$

$$\frac{E'_{x1}}{E_{x1}} = -\frac{H'_{y1}}{H_{y1}} = -1 \qquad [8]$$

Since H is magnetic field magnitude for the incident wave, the reflected wave is

$$H'_{y1} = He^{-j\beta_1 x \sin\theta}\, e^{j(\omega t + \beta_1 z \cos\theta)}$$

$$E'_{x1} = -\eta_1 H \cos\theta\, e^{-j\beta_1 x \sin\theta}\, e^{j(\omega t + \beta_1 z \cos\theta)}$$

$$E_{z1} = -\eta_1 H \sin\theta\, e^{-j\beta_1 x \sin\theta}\, e^{j(\omega t + \beta_1 z \cos\theta)}$$

And the resultant wave (incident and reflected) is

$$H_y = 2H \cos(\beta_1 z \cos\theta)\, e^{-j\beta_1 x \sin\theta}\, e^{j\omega t} \qquad [9]$$

$$E_x = -2j\eta_1 H \cos\theta \sin(\beta_1 z \cos\theta)\, e^{-j\beta_1 x \sin\theta}\, e^{j\omega t} \qquad [10]$$

$$E_z = -2\eta_1 H \sin\theta \cos(\beta_1 z \cos\theta)\, e^{-j\beta_1 x \sin\theta}\, e^{j\omega t} \qquad [11]$$

These are seen to be standing wave patterns with maxima and minima at half-wave points in front of the plane, *measured at the phase velocity in the direction normal to the plane.* That is, there are:

$$\text{Zeros of } E_x, \text{ maxima of } H_y \text{ and } E_z \text{ at } z = -\frac{n\lambda_1}{2\cos\theta}$$

$$\text{Zeros of } H_y \text{ and } E_z, \text{ maxima of } E_x \text{ at } z = -\frac{(2n+1)\lambda_1}{4\cos\theta}$$

where $\lambda_1 = 1/f\sqrt{\mu_1\epsilon_1}$.

B. Polarization Normal to the Plane of Incidence. The analysis is so similar to the previous case that all details need not be repeated.

The incident wave may be written

$$E_{y1} = Ee^{-j\beta_1 x \sin\theta}\, e^{j(\omega t - \beta_1 z \cos\theta)} \quad [12]$$

$$H_{x1} = -\frac{E}{\eta_1}\cos\theta\, e^{-j\beta_1 x \sin\theta}\, e^{j(\omega t - \beta_1 z \cos\theta)} \quad [13]$$

$$H_{z1} = \frac{E}{\eta_1}\sin\theta\, e^{-j\beta_1 x \sin\theta}\, e^{j(\omega t - \beta_1 z \cos\theta)} \quad [14]$$

The ratio $K = Z_2/Z_1$ is zero as before, so that

$$\frac{E'_{y1}}{E_{y1}} = -\frac{H'_{x1}}{H_{x1}} = -1 \quad [15]$$

The reflected wave is then

$$E'_{y1} = -Ee^{-j\beta_1 x \sin\theta}\, e^{j(\omega t + \beta_1 z \cos\theta)} \quad [16]$$

$$H'_{x1} = -\frac{E}{\eta_1}\cos\theta\, e^{-j\beta_1 x \sin\theta}\, e^{j(\omega t + \beta_1 z \cos\theta)} \quad [17]$$

$$H'_{z1} = -\frac{E}{\eta_1}\sin\theta\, e^{-j\beta_1 x \sin\theta}\, e^{j(\omega t + \beta_1 z \cos\theta)} \quad [18]$$

The resultant of incident plus reflected wave,

$$E_y = -2jE\sin(\beta_1 z\cos\theta)\, e^{-j\beta_1 x \sin\theta}\, e^{j\omega t} \quad [19]$$

$$H_x = -\frac{2E}{\eta_1}\cos\theta\cos(\beta_1 z\cos\theta)\, e^{-j\beta_1 x \sin\theta}\, e^{j\omega t} \quad [20]$$

$$H_z = -\frac{2jE}{\eta_1}\sin\theta\sin(\beta_1 z\cos\theta)\, e^{-j\beta_1 x \sin\theta}\, e^{j\omega t} \quad [21]$$

This is again a standing wave pattern in the z direction with:

Zeros of E_y and H_z, maxima of H_x at $z = -\dfrac{n\lambda_1}{2\cos\theta}$

Zeros of H_x, maxima of E_y and H_z at $z = -\dfrac{(2n+1)\lambda_1}{4\cos\theta}$

where $\lambda_1 = \dfrac{1}{f\sqrt{\mu_1\epsilon_1}}$.

It might be argued that the use of the impedance concept, with phase velocities in the direction normal to discontinuities, was perhaps no

easier for this simple case (at least so far as the mathematics is concerned) than a straightforward application of the boundary conditions. However, the point of view, the physical picture, is much superior. When it is grasped for simple cases such as this, it will prove extremely valuable for the analysis of more complex cases which will next be studied.

7.12 Incidence at any Angle on a Boundary between Perfect Dielectrics

For a wave crossing a boundary between perfect dielectrics the same division of the wave into two components employed in the previous article will be followed.

A. Polarization in Plane of Incidence. The incident wave is identical with that described by Eqs. 7.11(1), (2), and (3). There will be, just as for the case of normal incidence, some reflected wave and some wave transmitted to the second dielectric material. By the same reasoning used to determine the angle of reflection in the previous article, it follows that the phase velocities of each of these waves (incident, reflected and transmitted) tangential to the surface must be the same. That is,

$$\frac{v_1}{\sin\theta} = \frac{v_1}{\sin\theta'} = \frac{v_2}{\sin\theta''} \qquad [1]$$

where θ' = angle of reflection.

θ'' = angle of transmitted ray from normal (angle of refraction).

In addition to the fact that both the transmitted and reflected rays lie in the plane of incidence (no propagation in the y direction), it is now known from these conditions that the angle of reflection is equal to the angle of incidence, and that the angle of refraction is related to the angle of incidence by the following relation (known as Snell's law):

$$\frac{\sin\theta''}{\sin\theta} = \frac{v_2}{v_1} = \sqrt{\frac{\mu_1\epsilon_1}{\mu_2\epsilon_2}} \qquad [2]$$

The characteristic wave impedance for the incident wave in medium 1 is

$$Z_1 = \frac{E_{x1}}{H_{y1}} = \eta_1 \cos\theta$$

That for the reflected wave is just the negative of Z_1. The characteristic wave impedance for the wave in medium 2 is

$$Z_2 = \eta_2 \cos\theta''$$

The ratio of impedances,

$$K = \frac{Z_2}{Z_1} = \frac{\eta_2 \cos \theta''}{\eta_1 \cos \theta}$$

From (2), this may also be written

$$K = \frac{\eta_2 \sqrt{1 - \left(\dfrac{v_2}{v_1}\right)^2 \sin^2 \theta}}{\eta_1 \cos \theta} \tag{3}$$

Coefficients for the reflected and transmitted waves are given by Eqs. 7.07(7) and (8).

$$\frac{E'_{x1}}{E_{x1}} = -\frac{H'_{y1}}{H_{y1}} = \frac{K-1}{K+1} \tag{4}$$

$$\frac{E_{x2}}{E_{x1}} = K\frac{H_{y2}}{H_{y1}} = \frac{2K}{K+1} \tag{5}$$

From (3), (4), (5), and the incident wave, Eqs. 7.11(1), (2), and (3), transverse components of the reflected and transmitted waves are determined. The normal components are

$$E'_{z1} = E'_{x1} \tan \theta \tag{6}$$

$$E_{z2} = -E_{x2} \tan \theta'' \tag{7}$$

The resultant components of incident plus reflected wave in medium 1 may be found by adding the previous results. They are given below as ratios to the incident wave components. The expressions are written so as to make easy an interpretation of the traveling and standing wave parts for various values of K. Thus as K approaches unity the standing wave part disappears. As K approaches infinity or zero, the appropriate expression below shows the dominance of a particular standing wave over the vanishing traveling wave.

$$\frac{E_x}{E_{x1}} = 2\left[\frac{1}{K+1}e^{-j\beta' z} + \frac{K-1}{K+1}\cos \beta' z\right] \qquad (K > 1)$$

$$\frac{E_z}{E_{z1}} = \frac{H_y}{H_{y1}} = 2\left[\frac{1}{K+1}e^{-j\beta' z} - j\frac{K-1}{K+1}\sin \beta' z\right] \qquad (K > 1)$$

$$\frac{E_x}{E_{x1}} = 2\left[\frac{K}{K+1}e^{-j\beta' z} - j\frac{1-K}{1+K}\sin \beta' z\right] \qquad (K < 1)$$

$$\frac{E_z}{E_{z1}} = \frac{H_y}{H_{y1}} = 2\left[\frac{K}{K+1}e^{-j\beta' z} + \frac{1-K}{1+K}\cos \beta' z\right] \qquad (K < 1)$$

where again

$$K = \frac{\eta_2 \cos \theta''}{\eta_1 \cos \theta} = \frac{\eta_2}{\eta_1} \frac{\sqrt{1 - \left(\frac{v_2}{v_1}\right)^2 \sin^2 \theta}}{\cos \theta}$$

$$\beta' = \frac{2\pi \cos \theta}{\lambda_1} = \frac{2\pi f \cos \theta}{v_1}$$

B. Polarization Normal to Plane of Incidence. The incident wave is as in Eqs. 7.11(12), (13), and (14). The characteristic wave impedance for this incident wave is then

$$Z_1 = \frac{E_{y1}}{H_{x1}} = -\frac{\eta_1}{\cos \theta}$$

That for the reflected wave is the negative of Z_1. That for the transmitted wave is

$$Z_2 = \frac{E_{y2}}{H_{x2}} = -\frac{\eta_2}{\cos \theta''}$$

So the ratio of impedances is now

$$K = \frac{Z_2}{Z_1} = \frac{\eta_2 \cos \theta}{\eta_1 \cos \theta''} = \frac{\eta_2 \cos \theta}{\eta_1 \sqrt{1 - \left(\frac{v_2^2}{v_1^2}\right) \sin^2 \theta}} \tag{8}$$

From Eqs. 7.07(7) and (8)

$$\frac{E'_{y1}}{E_{y1}} = -\frac{H'_{x1}}{H_{x1}} = \frac{K - 1}{K + 1} \tag{9}$$

$$\frac{E_{y2}}{E_{y1}} = K \frac{H_{x2}}{H_{x1}} = \frac{2K}{K + 1} \tag{10}$$

From (8), (9), (10) and the incident wave, Eqs. 7.11(12) to (14), transverse components of the wave are determined. The normal components are

$$H'_{z1} = H'_{x1} \tan \theta$$

$$H_{z2} = -H_{x2} \tan \theta$$

The resultant components of incident plus reflected wave in medium 1

are again found by adding the previous results.

$$\frac{H_x}{H_{x1}} = 2\left[\frac{1}{K+1}e^{-j\beta' z} - j\frac{K-1}{K+1}\sin\beta' z\right] \qquad (K > 1)$$

$$\frac{H_z}{H_{z1}} = \frac{E_y}{E_{y1}} = 2\left[\frac{1}{K+1}e^{-j\beta' z} + \frac{K-1}{K+1}\cos\beta' z\right] \qquad (K > 1)$$

$$\frac{H_x}{H_{x1}} = 2\left[\frac{K}{K+1}e^{-j\beta' z} + \frac{1-K}{1+K}\cos\beta' z\right] \qquad (K < 1)$$

$$\frac{H_z}{H_{z1}} = \frac{E_y}{E_{y1}} = 2\left[\frac{K}{K+1}e^{-j\beta' z} - j\frac{1-K}{1+K}\cos\beta' z\right] \qquad (K < 1)$$

and again

$$K = \frac{\eta_2 \cos\theta}{\eta_1 \cos\theta''} = \frac{\eta_2\cos\theta}{\eta_1\sqrt{1 - \left(\frac{v_2}{v_1}\right)^2 \sin^2\theta}}$$

$$\beta' = \frac{2\pi\cos\theta}{\lambda_1} = \frac{2\pi f\cos\theta}{v_1}$$

Problem 7.12. By the concepts of wave propagation for general angles of incidence developed in Art. 7.12, extend the analysis of the quarter-wave matching section for eliminating reflection between dielectrics developed in Art. 7.09 for normal incidence. That is, without resorting to matching of wave solutions, determine the thickness of a matching section and its dielectric constant and permeability if it is perfectly to eliminate reflections for a single-frequency wave incident at the angle θ upon a plane surface separating two dielectrics of constants ϵ_1, μ_1 and ϵ_2, μ_2, respectively.

7.13 Total Reflection

A study of the general results from the previous article for incidence on a dielectric at any angle shows that there are several particular angles of incidence that are of special interest. First, under what conditions might a wave be totally reflected? Previous study has shown that this occurs when there is complete mismatch, that is, when the ratio of wave impedances K is either zero or infinity. For a wave polarized in the plane of incidence, this ratio is given by Eq. 7.12(3).

$$K = \frac{\eta_2}{\eta_1}\frac{\sqrt{1 - \left(\frac{v_2}{v_1}\right)^2 \sin^2\theta}}{\cos\theta} \qquad [1]$$

This factor may be made infinite if $\cos\theta = 0$, but this is a trivial case since it represents a wave traveling parallel to the boundary. The

factor may become zero at a critical angle such that

$$1 - \frac{v_2^2}{v_1^2}\sin^2\theta_c = 0$$

$$\sin\theta_c = \frac{v_1}{v_2} = \sqrt{\frac{\mu_2\epsilon_2}{\mu_1\epsilon_1}} \tag{2}$$

Before investigating this factor further, let us notice the impedance ratio K for waves polarized normal to the plane of incidence. Equation 7.12(8) gives this:

$$K = \frac{\eta_2}{\eta_1}\frac{\cos\theta}{\sqrt{1 - \left(\frac{v_2}{v_1}\right)^2\sin^2\theta}}$$

This factor becomes zero for waves traveling parallel to the boundary and infinite for waves incident at an angle defined by (2). It is then apparent that any type of wave incident at an angle which satisfies (2) is totally reflected. A study of the refracted wave makes it clear why this should be.

Consider, for ease of visualization, two dielectrics of the same permeability. Equation (2) then reduces to

$$\sin\theta_c = \sqrt{\frac{\epsilon_2}{\epsilon_1}} \tag{3}$$

A real solution for θ_c exists only if $\epsilon_2 \le \epsilon_1$, which means that the wave must be passing from one dielectric to another of smaller dielectric constant. Equation 7.12(2) shows that the refracted wave in the second medium then makes a larger angle with the normal than the incident wave. In particular, it is possible for the transmitted or refracted wave to make an angle of 90° with the normal (that is, become tangential to the surface) when the incident wave is still at some angle θ_c less than 90° from the normal. For angles of incidence greater than θ_c, the angle of refraction could not be determined since (5) would require that it have a sine greater than unity. This means, of course, that there is total reflection over all the range of incident angles between the critical angle, defined by (2), and $\theta = 90°$. This is apparent also from the fact that imaginary values of the impedance ratios result over all this range.

When we say that no wave is transmitted across the boundary for imaginary values of K, it should not be assumed that there is absolutely no field in the region beyond the boundary. It will be noted that when K is imaginary, the usual expression can still be written with complete correctness for the fields beyond the boundary. These expressions will

have an exponential factor which instead of being of the form $e^{j\beta z}$ where β is a real quantity will be of the form $e^{-\alpha z}$ where α is real. In other words, it is true that no *propagating waves* extend beyond the boundary; the field *does* penetrate into the second dielectric, but it dies off exponentially.

Since K is imaginary for angles of incidence greater than the critical, an inspection of the Eqs. 7.12(4) and 7.12(9) shows that although the reflected wave always has a magnitude equal to the incident wave, the phase angle can take on various values. In fact if ψ is the phase of E'_{x1}/E_{x1} and ψ' is the phase of E'_{y1}/E_{y1}, the phase angles may be obtained from Eqs. 7.12(4) and 7.12(9) and the definitions of K.

$$\tan \psi = \frac{-2\frac{\eta_2}{\eta_1}\cos\theta\sqrt{\left(\frac{v_2}{v_1}\right)^2 \sin^2\theta - 1}}{\cos^2\theta - \left(\frac{\eta_2}{\eta_1}\right)^2\left(\frac{v_2^2}{v_1^2}\sin^2\theta - 1\right)} \quad \text{polarization in incident plane} \qquad [4]$$

$$\tan \psi' = \frac{-2\frac{\eta_2}{\eta_1}\cos\theta\sqrt{\left(\frac{v_2}{v_1}\right)^2 \sin^2\theta - 1}}{\left(\frac{\eta_2}{\eta_1}\right)^2\cos^2\theta - \left(\frac{v_2^2}{v_1^2}\sin^2\theta - 1\right)} \quad \text{polarization normal to incident plane} \qquad [5]$$

The x component thus has a different phase angle from the y component for angles of incidence giving total reflection. So, if wave components of both polarization are present in the incident wave, the reflected wave under these conditions will be elliptically polarized (Art. 7.05).

Problem 7.13(*a*). Calculate the critical angle for an electromagnetic wave passing from the following dielectrics into air.

Material	ϵ/ϵ_0 (ratio of dielectric constant to that of air)
Distilled water	81.1
Ethyl alcohol	25.8
Glass (high density)	9
Glass (low density)	6
Mica	6
Quartz	5
Petroleum oil	2.1

Problem 7.13(*b*). Show that the phase difference between the two polarization components in the reflected wave under conditions of total reflection, $\delta = \psi - \psi'$ is given by

$$\tan\left(\frac{\delta}{2}\right) = \frac{\frac{\eta_2}{\eta_1}\left[\left(\frac{v_2}{v_1}\right)^2 - 1\right]\sin^2\theta}{\left[\left(\frac{\eta_2}{\eta_1}\right)^2 - 1\right]\cos\theta\sqrt{\left(\frac{v_2}{v_1}\right)^2\sin^2\theta - 1}}$$

Problem 7.13(c). Find the expressions for the fields in the second dielectric when the incident angle is such as to yield imaginary K in such form as to disclose the exponential decay of these fields with penetration into the second dielectric.

7.14 Polarizing Angle

Let us next ask under what conditions there may be no reflected wave. That is, under what conditions may the two wave impedances match exactly, making K unity, without resorting to an intervening dielectric matching section? For a wave polarized in the plane of incidence, this requires setting Eq. 7.13(1) equal to unity.

$$K = \frac{\eta_2}{\eta_1}\,\frac{\sqrt{1 - \left(\dfrac{v_2}{v_1}\right)^2 \sin^2\theta}}{\cos\theta} = 1$$

The solution of this equation results in

$$\sin\theta = \sqrt{\frac{1 - \left(\dfrac{\mu_2\epsilon_1}{\mu_1\epsilon_2}\right)}{1 - \left(\dfrac{\epsilon_1}{\epsilon_2}\right)^2}} \qquad [1]$$

Similarly, the impedance ratio for waves polarized normal to the plane of incidence may be made unity if

$$\sin\theta = \sqrt{\frac{1 - \dfrac{\mu_1\epsilon_2}{\mu_2\epsilon_1}}{1 - \left(\dfrac{\mu_1}{\mu_2}\right)^2}} \qquad [2]$$

For angles of incidence defined by (1) and (2) there would be no reflected wave. Let us see what numerical values these angles might have in practical cases. There are few if any dielectrics of practical importance in radio applications that have permeabilities substantially different from that of air. If μ_1 is made equal to μ_2 in the above equations, it is seen that (2) can have no real solutions (since $\sin\theta = \infty$) unless $\epsilon_1 = \epsilon_2$ as well, which is the trivial case of identical permeabilities and dielectric constants. However, (1) may be solved to yield an angle $\theta = \theta_p$ for which there is no reflected wave when the incident wave is polarized in the plane of incidence.

$$\sin\theta_p = \frac{1}{\sqrt{1 + \dfrac{\epsilon_1}{\epsilon_2}}}$$

$$\tan\theta_p = \sqrt{\frac{\epsilon_2}{\epsilon_1}} \qquad [3]$$

(If materials commonly were to have identical dielectric constants and different permeabilities, (2) might conversely have a solution whereas (1) would have none.)

Although a wave polarized in the plane of incidence and incident at the angle defined by (3) will have no reflected component, a wave polarized normal to the plane of incidence and incident at this angle will have reflected components. Consequently if a wave containing both components is incident at this angle, one component causes reflections while the other produces no reflections. The reflected wave is then polarized normal to the plane of incidence, even though the incident wave was unpolarized. The angle θ_p defined by (3), or (1) in general, is consequently known as the polarizing angle. It is also often known as the Brewster angle.

Notice that K is less than unity for angles of incidence $\theta < \theta_p$ and greater than unity for angles of incidence $\theta > \theta_p$. An inspection of the equations of Art. 7.12 shows that the standing wave patterns change form at this angle. Thus in the standing wave pattern there is a minimum of E_x and maxima of E_z and H_y at the dielectric surface for angles $\theta < \theta_p$. For angles $\theta > \theta_p$ there is a maximum of E_x and minima of E_z and H_y at the boundary. Of course exactly at the polarizing angle the standing wave components of E_x, E_z, and H_y disappear completely.

Problem 7.14. For the dielectrics listed in Prob. 7.13(a), determine the polarizing angle for waves passing from each of the dielectrics into air, and also for waves passing from air into the dielectrics.

WAVES IN IMPERFECT CONDUCTORS AND DIELECTRICS

7.15 Waves in Conducting Materials

From Poynting's theorem it is known that no energy can be transmitted into a perfect conductor, and so no wave can exist inside such a conductor. Furthermore, no fields of any kind, waves or otherwise, can be in such a conductor. If the conductivity is not perfect, electric and magnetic fields may exist inside the conductor, as was shown in the past chapter, and under certain conditions it may be desirable to consider these as waves.

For a conductor, the equations corresponding to Eq. 7.02(3) and Eq. 7.02(4), assuming sinusoidal time variations, are

$$\nabla \times \bar{E} = -j\omega\mu\bar{H}$$

$$\nabla \times \bar{H} = (\sigma + j\omega\epsilon)\bar{E} = j\omega\epsilon\left[1 + \frac{\sigma}{j\omega\epsilon}\right]\bar{E}$$

It is apparent from these equations that all mathematical manipulations

of previous sections are valid if

$$\epsilon_c = \epsilon\left[1 + \frac{\sigma}{j\omega\epsilon}\right] \tag{1}$$

is substituted in place of ϵ for solutions applying inside the conducting material. In other words, as far as the use of previously derived mathematical relations are concerned, a conductor is simply another dielectric with a complex dielectric constant ϵ_c and with its conductivity never appearing explicitly. Of course, we are interested in more than the mathematical relations so we shall return soon to see what this means physically.

Thus taking γ again as the propagation constant for a uniform plane wave, Eq. 7.04(3),

$$\gamma = j\omega\sqrt{\mu\epsilon_c}$$

Since ϵ_c is complex, γ will have real and imaginary parts. Thus

$$\gamma = \alpha + j\beta = j\omega\sqrt{\mu\epsilon\left(1 + \frac{\sigma}{j\omega\epsilon}\right)} \tag{2}$$

$$\alpha = \omega\sqrt{\frac{\mu\epsilon}{2}\left(\sqrt{1 + \frac{\sigma^2}{\omega^2\epsilon^2}} - 1\right)} \tag{3}$$

$$\beta = \omega\sqrt{\frac{\mu\epsilon}{2}\left(\sqrt{1 + \frac{\sigma^2}{\omega^2\epsilon^2}} + 1\right)} \tag{4}$$

The intrinsic impedance for the conducting material

$$\eta_c = \sqrt{\frac{\mu}{\epsilon_c}} = \sqrt{\frac{\mu}{\epsilon\left[1 + \frac{\sigma}{j\omega\epsilon}\right]}} \tag{5}$$

These values may be substituted in all previous general wave results. It is apparent at once that since γ has real and imaginary parts, there is now attenuation as the wave progresses in the conductor

$$e^{-\gamma z} = e^{-j\beta z}e^{-\alpha z}$$

This is as would be expected, since energy is lost by currents flowing in the imperfect conductor. Since η_c is complex, it follows that electric and magnetic fields are not in phase for a uniform plane wave in a conductor as they were in a perfect dielectric. Since K, the impedance ratio between η_c and the intrinsic impedance of a dielectric, will also be complex, any waves reflected by passage from a dielectric to a conducting

medium will have phase differences with respect to the incident wave other than the 0° or 180° values found for reflection from perfect dielectrics and conductors.

The special cases of greatest interest are those in which the material is either a reasonably good conductor, or a reasonably good dielectric, and of these more detailed analysis will follow.

7.16 Waves in Imperfect Conductors

An imperfect conductor will be regarded as a conductor in which displacement currents are negligibly small compared with conduction currents for the frequency of interest but in which the resistivity cannot be neglected. That is,

$$\frac{\sigma}{\ } \gg 1$$

Then Eq. 7.15(2) reduces to

$$\gamma = j\omega\sqrt{\frac{\mu\sigma}{j\omega}} = (1 + j)\sqrt{\pi f \mu\sigma} = \frac{1 + j}{\delta} \qquad [1]$$

δ is the depth of penetration used extensively in Chapter 6 and defined by Eq. 6.04(10). The propagation function for the wave,

$$e^{-\gamma z} = e^{-\frac{z}{\delta}} e^{-j\frac{z}{\delta}}$$

shows that the wave decreases in magnitude exponentially, and has decreased to $1/e$ of its original value after propagating a distance equal to depth of penetration of the material. The phase factor corresponds to a very small phase velocity,

$$v = \frac{\omega}{\beta} = \omega\delta = c\frac{2\pi\delta}{\lambda_0} \qquad [2]$$

where c = velocity of light in free space.
λ_0 = free space wavelength.

Since δ/λ_0 is usually very small (see Fig. 6.05*a*) this phase velocity is usually much less than the velocity of light.

Equation 7.15(5) gives, for a good conductor,

$$\eta_c = \sqrt{\frac{j\omega\mu}{\sigma}} = (1 + j)\sqrt{\frac{\pi f \mu}{\sigma}} = (1 + j)R_s \qquad [3]$$

R_s is the surface resistivity or high-frequency skin effect resistance per square of a plane conductor of great depth. Equation (3) shows that

electric and magnetic fields are 45° out of time phase for the wave propagating in a good conductor. Also, since R_s is very small (see Fig. 6.05a) the ratio of electric field to magnetic field in the wave is small.

Since R_s is much less than unity for ordinary conducting materials (0.014 for copper at 3000 mc) and since the intrinsic impedance of most dielectrics is much greater than unity (377 for air) the ratio K which appears in the reflection formulas for waves incident upon conducting boundaries will be very small.

$$K = \frac{Z_2}{Z_1}$$

$$= \frac{\eta_c}{\eta_1} \text{ for normal incidence}$$

$$= \frac{(1 + j)R_s}{\eta_1}$$

Reflection from such conductors is then, for most practical cases, accurately enough computed by the results found for a perfect conductor, $K = 0$. A small amount of energy, of course, is transferred to the imperfect conductor to take care of losses due to the current flow in it, and a small 90° out-of-phase component is reflected from the surface of the conductor.

The above results make it clear that this wave picture presents another way of looking at skin effect phenomena, as was predicted in Art. 6.02. The decrease in current density and field strengths as one progresses into the conductor may be thought of as the attenuation due to the finite conductivity, which corresponds to distributed conductance in a transmission line.

Problem 7.16(a). Compute the percentage of energy transmitted to a reasonably good conductor when a plane uniform wave is normally incident upon it.

Problem 7.16(b). Compute the ratio of 90° out-of-phase component to the in-phase component in the reflected wave caused by a normally incident wave upon a conductor.

Problem 7.16(c). Calculate values for the results of a and b for incidence from air to copper at 30 mc/sec and at 3000 mc/sec.

7.17 Imperfect Dielectrics

In a dielectric material with finite conductivity, it is not wise to neglect displacement currents as was done for good conductors, since displacement currents will usually be much greater than conduction currents if the material is to be useful as a dielectric. Neither can we completely

neglect conductive currents if any information is to be obtained on the effect of losses. It seems necessary to consider both the conductivity and dielectric constant terms in the expressions of Art. 7.15. That is, the complex dielectric constant is

$$\epsilon_c = \epsilon\left[1 + \frac{\sigma}{j\omega\epsilon}\right] \quad [1]$$

The properties of a lossy dielectric might be expressed by stating σ and ϵ. However, for reasons having to do with measurement and variation of properties with frequency, it is more common to express the properties of a dielectric in terms of two quantities, ϵ' and ϵ'', such that

$$\epsilon_c = \epsilon_0[\epsilon' - j\epsilon''] \quad [2]$$

ϵ_0 is the dielectric constant of free space in mks units, ϵ' is the familiar value of dielectric constant for the material, based on air or space as unity, and ϵ'' is called the loss factor. By comparing (1) and (2) we see that

$$\epsilon'' = \frac{\sigma}{\omega\epsilon_0} = \frac{36\pi\sigma}{\omega \times 10^{-9}} \quad [3]$$

where σ is in mhos per meter.

The ratio of ϵ''/ϵ' is also a common constant for dielectrics, since it is a direct measure of the ratio of conduction current to displacement current in the dielectric.

$$\frac{\epsilon''}{\epsilon'} = \frac{\sigma}{\omega\epsilon_0\epsilon'} = \frac{\sigma}{\omega\epsilon} \quad [4]$$

This ratio is often called power factor of the dielectric, although it is only an approximation to power factor good for small values of ϵ''/ϵ'. Strictly, power factor is defined as

$$\text{P.F.} = \sin\phi$$

where

$$\tan\phi = \frac{\epsilon''}{\epsilon'} \quad [5]$$

However, since for most useful dielectrics the ratio of conduction to displacement current is less than 0.10, it is satisfactory to use the ratio (4) as power factor.

It should be emphasized that no matter what quantities are used to express the properties of a dielectric, σ, ϵ, ϵ', ϵ'', or P.F., any of these may in general be a function of frequency. Unlike the situation for conduc-

tors, where a value of conductivity σ will hold for all frequencies of interest, properties of certain dielectrics given for one frequency may not indicate at all the properties at another frequency.

7.18 Waves in Imperfect Dielectrics

It will be assumed here that the dielectric is good enough so that conduction currents are relatively small compared to displacement currents. That is, power factor is assumed to be small. With this assumption,

$$\frac{\epsilon''}{\epsilon'} = \frac{\sigma}{\omega\epsilon} \ll 1$$

Eqs. 7.15(3) and 7.15(4) reduce to

$$\alpha \cong \frac{k\epsilon''}{2\epsilon'} \qquad [1]$$

$$\beta \cong k\left[1 + \frac{1}{8}\left(\frac{\epsilon''}{\epsilon'}\right)^2\right] \qquad [2]$$

where $k = \omega\sqrt{\mu\epsilon} = 2\pi/\lambda$.

Thus the exponential expressing attenuation is

$$e^{-\alpha z} = e^{-\frac{k\epsilon'' z}{2\epsilon'}} = e^{-\frac{\epsilon''\pi}{\epsilon'\lambda}z}$$

We see that the wave has attenuated to $1/e$ (about 37 per cent) of its original value in a distance

$$z = \frac{\lambda\epsilon'}{\epsilon''\pi}$$

If $\epsilon''\pi/\epsilon'$ is small compared with unity, this distance is large compared to wavelength.

From (2) the phase velocity is

$$v = \frac{\omega}{\beta} = \frac{1}{\sqrt{\mu\epsilon}\left[1 + \frac{1}{8}\left(\frac{\epsilon''}{\epsilon'}\right)^2\right]} \qquad [3]$$

The phase velocity is thus decreased a small amount by the conductivity of the dielectric.

The intrinsic impedance of the medium is given by Eq. 7.15(5).

$$\eta_c = \sqrt{\frac{\mu}{\epsilon\left(1 - j\frac{\epsilon''}{\epsilon'}\right)}} \cong \eta\left\{\left[1 + \frac{3}{8}\left(\frac{\epsilon''}{\epsilon'}\right)^2\right] + j\frac{\epsilon''}{2\epsilon'}\right\} \qquad [4]$$

This expression shows a small 90° out-of-phase component between electric and magnetic fields in the wave propagating in an imperfect dielectric. The ratio between in-phase components is also changed by the small correction factor $\frac{3}{8}(\epsilon''/\epsilon')^2$.

If the above complex value of η_c is substituted in the impedance ratio K in order to determine reflections from the imperfect conductor, K will be found to be complex also. This means, of course, that there will be a slight effect of the conductivity on the magnitudes of the reflected and transmitted components, and that a small phase angle will exist between these components and the incident wave. Since this phase angle need not be the same for both polarization components, the reflected wave may be elliptically polarized if the incident wave contains components polarized in both planes.

7.19 Properties and Classification of Poor Conductors

Special results have been given for reasonably good conductors and reasonably good dielectrics; we wish to know to what materials these conclusions may be applied. Practically any metallic conductor has a value of $\sigma/\omega\epsilon$ much greater than unity at any radio frequency, so that these may be considered reasonably good conductors. Most dielectrics or insulating materials in practical use as dielectrics have values of $\sigma/\omega\epsilon$ (power factor) which are not large compared with unity (say less than 0.05), so that the results of Art. 7.18 may be applied to these. Some other materials of importance to radio may be considered as reasonably good conductors at some radio frequencies and reasonably good dielectrics at other frequencies. Such materials are earth and sea water.

Below are listed for certain of these materials the frequency at which $\sigma/\omega\epsilon = 100$. For all frequencies below this, $\sigma/\omega\epsilon$ should be very large compared with unity so that the results of Art. 7.16 may be applied. Also there is tabulated the frequency at which $\sigma/\omega\epsilon = 0.05$. For all frequencies above this the approximations of Art. 7.18 should be accurate enough. For any frequencies between these limits, the exact expressions of Art. 7.15 should be used if accurate results are desired. It will be recognized that the following values are merely representative since these materials vary greatly in electrical properties.

Material	σ Conductivity Mhos/Meter	ϵ'	Frequency at Which $\sigma/\omega\epsilon = 100$ (Art 7.16 valid for all lower frequencies)	Frequency at Which $\sigma/\omega\epsilon = 0.05$ (Art. 7.18 valid for all higher frequencies)
Sea water	4	81	8.9×10^6	$17{,}800 \times 10^6$
Fresh water	10^{-3}	81	2.2×10^3	4.4×10^6
Wet earth	10^{-3}	10	18.0×10^3	36×10^6
Dry earth	10^{-5}	5	0.36×10^3	0.72×10^6

7.20 Elimination of Wave Reflections for Incidence on Good Conductors

For high-frequency applications it is often desirable to reduce or eliminate spurious reflections from metallic objects placed in the vicinity of radiating systems. We shall show[3] that a thin conducting film may be utilized for this purpose if removed a quarter wavelength from the metallic surface. This example will again serve to illustrate the usefulness of the transmission line analogy and impedance concepts.

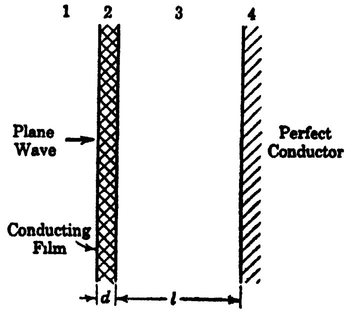


FIG. 7.20a. Impedance sheet for termination of a wave region.

The uniform plane wave normally incident upon a good conductor, (4) of Fig. 7.20a, will be considered. It has been shown (Art. 7.06) that a standing wave pattern is set up due to the combination of reflected and incident waves, so that a quarter wavelength in front of the conductor there is a minimum of magnetic field and a maximum of electric field. This represents a point of very high impedance, E/H. Suppose a given thickness, d, of any material is placed at that point. The impedance viewed from the front surface of the material, where the wave strikes, may be expressed in terms of the terminating impedance, the thickness, and the propagation constant through that material, Eq. 7.07(10). If the back surface of the film is placed exactly at the node of magnetic field, the terminating impedance is practically infinite. (It is of course exactly infinite if the conductor 4 is perfect.) The impedance at the front surface is then

$$Z_i = \eta_2 \coth \gamma_2 d$$

[3] S. A. Schelkunoff, "The Electromagnetic Theory of Coaxial Transmission Lines and Cylindrical Shields," *Bell System Tech. Journ.*, 13, 532 (October, 1934).

For perfect matching and complete elimination of reflections, this impedance should be equal to the characteristic wave impedance for the wave in the dielectric material 1. For a uniform plane wave this is merely

$$\eta_1 = \eta_2 \coth \gamma_2 d$$

Let us try to accomplish this matching with a film of conductivity σ, and of small enough thickness so that

$$\coth \gamma_2 d \cong \frac{1}{\gamma_2 d}$$

The values of γ_2 and η_2 for a conducting material may be substituted from Art. 7.15.

$$\eta_1 \cong \left[\sqrt{\frac{\mu_2}{\epsilon_2\left(1 + \dfrac{\sigma_2}{j\omega\epsilon_2}\right)}}\right]\left[\frac{1}{j\omega d\sqrt{\mu\epsilon_2\left(1 + \dfrac{\sigma_2}{j\omega\epsilon_2}\right)}}\right]$$

$$\eta_1 = \frac{1}{j\omega\epsilon_2 d\left(1 + \dfrac{\sigma_2}{j\omega\epsilon_2}\right)}$$

The impedance η_1 is purely real. The right side may only be a pure real number for a reasonably good conductor such that unity is negligible compared with $\sigma/\omega\epsilon$. Then

$$\eta_1 = \frac{1}{\sigma_2 d} \qquad [1]$$

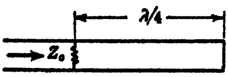


FIG. 7.20*b*. Transmission line equivalent of Fig. 7.20*a*.

This corresponds to a thin film of resistive material whose resistance per square is equal to the intrinsic impedance η_1. This is analogous in transmission line terms to the characteristic impedance of a transmission line placed a quarter wavelength in front of a short-circuited end (Fig. 7.20*b*). Since the short-circuited quarter-wave line has infinite impedance, this represents perfect matching for a wave approaching from the left.

Note that the conductivity σ_2 must be quite small if d is not to be unreasonably small in thickness. Thus if the material 1 is air or space ($\eta = 120\pi$ ohms), σ_2 must be 28.5 mhos per meter to make $d = 0.1$ mm. This corresponds to a conductivity about 0.5×10^{-6} times that of copper. Note also that the spacing l between film and conductor is a

quarter wavelength in the dielectric material of 3, so that this spacing may be decreased if a material of higher dielectric constant is used.

$$l = \frac{\lambda_3}{4} = \frac{1}{4f\sqrt{\mu_3\epsilon_3}} \qquad [2]$$

The perfect matching was possible because the film, which must absorb the incident wave, was placed a quarter wavelength from the conductor where the electric field was high. Matching is not possible with a film of simple electrical properties if it is attempted to place the film on the surface of the conductor itself, since this is a region of low electric field. The dielectric and conductivity properties of the film would then be unimportant.

Problem 7.20(a). Show that it is not possible to match the wave impedance exactly with any type of conducting film placed directly on the surface of the metal so that its terminating impedance is effectively zero.

Problem 7.20(b). By use of the transmission line analogies, determine the spacing between a film and a good conductor, and the conductivity properties of that film if reflections are to be perfectly eliminated for a wave incident at an angle θ from the normal.

8

GUIDED ELECTROMAGNETIC WAVES

8.01 Introduction

In the preceding chapter we were interested primarily in electromagnetic waves in boundless dielectrics except in so far as reflecting discontinuities were concerned. Now we wish to study specifically the behavior of these waves in the immediate vicinity of conducting and dielectric boundaries when the configurations of these boundaries have the effect of guiding the energy along their surfaces.

Actually no wave is ever truly free from the effect of conductors and dielectrics, although for radio waves in the space between transmitter and receiver and at a great distance from the ground, the effects of the boundaries may be negligible for all practical purposes. Moreover, the problem of reflection of electromagnetic waves from a conductor or dielectric is not necessarily a different problem from that to be considered here, for it was shown in the past chapter that a uniform plane wave striking a plane discontinuity at a given angle of incidence may be looked upon as a non-uniform wave propagating parallel to the discontinuity. It might be said then that this non-uniform wave was being "guided" parallel to the boundary. No matter what is said, such a wave will also have a component of power flow (Poynting vector) normal to the boundary, except in certain special cases, so that one cannot think freely in terms of straightforward guiding or constraining of the energy to travel *entirely along* the boundary.

A viewpoint then for guided waves will be that of guiding electromagnetic energy primarily along the direction of the guiding system. Actually, if there is a transmission line or other form of guide, there must be some component of energy flow from the wave into the guiding boundary to account for that energy dissipated in the finite conductivity of the conductor or dielectric. Nevertheless, the principal energy flow (if the line is any good for energy transfer) is along the direction of the line. There is certainly no large component of energy flow from the line out into space under normal conditions.

The term guided energy has another implication. It seems to say that the wave is definitely following the guide, so that the guide could be taken around corners or tied into knots, and the wave would faith-

fully follow the path prescribed by the guide. We know that for ordinary transmission lines this is true, at least within certain reasonable limits. The transmission line certainly need not take an absolutely straight path between the source and its load to insure an energy transfer.

The qualification " within certain reasonable limits " was introduced above, for if the discontinuities are too abrupt, not all the wave will follow the line; some will be reflected and some may actually proceed into space, and this is called radiation. The problem of how much of a wave will continue along the conductor and how much will pass into space is a fairly tough one, requiring special techniques for handling. These will be developed in the separate chapter devoted to radiating systems.

It might at first seem queer that any of the wave should follow the guiding conductor if it were suddenly to make a sharp bend and go off in a direction entirely different from that in which the wave was previously proceeding. If the first point of view were the only one, merely that of a wave propagating energy in the direction of the guide, it might seem only a coincidence that this wave and the guide should have the same direction at any particular place. Actually this is not the complete story, for the two are intimately connected through the current flow in the guide and the charges on the guide arising from the magnetic and electric fields of this electromagnetic wave. If the conductor changes direction, the current flow will not ordinarily be expected to stop obligingly at that point, but rather tends to follow the conductor. This current flow in the new direction in turn generates a new wave in that direction, either of the same kind or of a different kind from that which existed before the bend.

Many of these concepts will become clearer as specific examples are discussed. The mathematical basis is of course not different from that of the past chapter. That is, solutions to the wave equation are required which fit the boundary conditions imposed by the conducting and dielectric guides. However, we shall be interested here in those solutions which represent energy transfer along the direction of the guide and which are intimately tied to the guide through some condition of current flow, charge induction, or special reflection. Analysis will then be confined mostly to guides which are straight and uniform, assuming that the wave will follow these well enough if there are reasonable changes in direction without worrying too much about the exact mechanism by which it occurs in any particular case. Guiding systems in which an important part of the energy does leave the guide at the discontinuities must patiently await consideration until the analysis of radiating systems is studied.

SIMPLE EXAMPLES OF GUIDED WAVES AND WAVE GUIDES

8.02 Waves Guided by an Infinite, Perfectly Conducting Plane

The concepts of the last chapter permit us to consider any plane wave incident upon a plane boundary as a non-uniform wave propagating parallel to the boundary. Consequently many waves might be considered as propagating parallel to the surface of a plane perfect conductor, corresponding to waves incident upon this plane from a perfect dielectric at different angles of incidence. Components for these waves are obtainable from Art. 7.11. However, for present purposes interest is only in waves for which energy transfer is primarily in the direction of the guide, so that the Poynting vector should be parallel to the plane. The only conditions under which there is never a component of Poynting vector normal to the plane are those which make the components of $\bar{E}$ and $\bar{H}$ zero in the direction of propagation parallel to the plane.

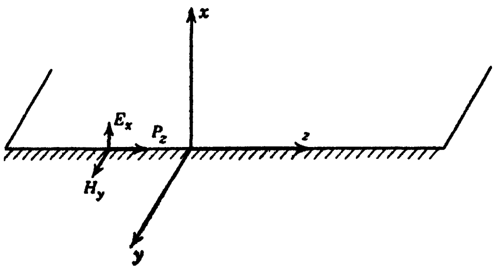


FIG. 8.02. Electromagnetic field vectors for a wave guided by a conducting plane.

Let us this time take the surface of the plane perfect conductor as lying in the Y–Z coordinate plane (Fig. 8.02). The results of Art. 7.11 cannot then be used directly because of the different orientation of the coordinate system, but the components for the new orientation are easily obtained, and it is desirable to keep the z direction as the direction of propagation. For a wave polarized in the plane of incidence, there is now a component of H_y. There would then also be a value of

$$P_x = (\bar{E} \times \bar{H})_x$$

at some instant of time unless E_z were zero. Similarly, there is a value of E_y for the wave polarized normal to the plane of incidence, so that there would be a value of P_x unless H_z were zero. A wave polarized in the plane of incidence may exist with $E_z = 0$ only when it is propagating exactly parallel to the boundary. The remaining components are then

$$H_y = He^{j(\omega t - k_1 z)} \qquad [1]$$

$$E_x = \eta_1 He^{j(\omega t - k_1 z)} \qquad [2]$$

where

$$k_1 = \omega\sqrt{\mu_1\epsilon_1} = \frac{\omega}{v_1}; \quad \eta_1 = \sqrt{\frac{\mu_1}{\epsilon_1}}$$

The subscripts 1 are to be used here and throughout the chapter to denote constants applying to the dielectric.

A wave polarized normal to the plane of incidence cannot exist under the required conditions of zero P_x. To make $H_z = 0$ requires propagation parallel to the boundary as before; but in this case all field components reduce to zero by the requirement of tangential electric field (E_y) zero at the conductor and therefore everywhere, since there are no variations with x if propagation is exactly parallel to the conductor.

The components of the wave defined by (1) and (2) might have been set down at once, without reference to results of the past chapter. It seems fairly evident that a uniform plane wave propagating parallel to the plane perfect conductor should get along all right if its electric field is oriented normal to the plane, as is required by the boundary condition imposed by the perfect conductor. There are, of course, charges induced on the conducting plane from the normal electric field component which ends on it. Also, the magnetic field component, normal to the electric field and therefore parallel to the plane, is related to the electric field by η_1, the intrinsic impedance of the dielectric. It produces a current flow in the plane. The amount and direction of this current flow (Art. 6.07) are given by

$$J = \bar{n} \times \bar{H}$$

So that current flow is entirely in the direction of propagation and has a magnitude in current per unit width equal to the magnetic field

$$J_z = H_y \tag{3}$$

Thus we see that this wave satisfies a second requirement to be considered as a guided wave. It is intimately tied to the conducting plane through the induced charges and currents. If the plane were to make a bend, there would at least be a tendency for this current to follow the plane, inducing a wave traveling in the new direction.

Before leaving this wave, a more critical examination might be made of the reasons for choosing it in preference to all other possibilities. Although it was stated that this wave was selected because it had no energy flow normal to the plane at any time, no other wave which might have been chosen could have any *average* energy flow normal to the plane. That is, on the average, there is as much energy traveling away

from the plane in the reflected wave as there is traveling toward the plane in the incident wave, for certainly no energy can penetrate the perfect conductor. This is further evident in that the components of electric and magnetic field producing the Poynting vector component normal to the surface are 90° out of time phase for all waves of Art. 7.11, striking a conducting surface at any general angle of incidence. Consequently no time average value of Poynting vector normal to the plane exists for any of the waves, although there may be a non-zero value at any instant, representing first a net power flow away from, and then later a net power flow toward the plane. Thus waves which are incident at a general angle θ and which have a definite standing wave pattern in the xy plane could be considered as guided by the plane with a phase velocity $v_1/\cos\theta$ in the z direction. It is all a matter of point of view, and the most convenient point of view is usually determined by the mechanism by which the waves are excited. This has not yet been studied at all.

8.03 Approximate Characteristics of a Wave Guided by an Imperfectly Conducting Plane

The plane wave of Art. 8.02, guided by the perfectly conducting plane, was found to propagate parallel to the conductor with the velocity of light in the dielectric. The wave has an electric field component normal to the plane which is η_1 times the magnetic field component parallel to the plane. These field components produce charges on the conductor and current flow in the direction of propagation. If the plane is imperfectly conducting, these conditions will be disturbed to some extent. This case may be solved exactly to determine the effect of the imperfect conductivity. However, only for such simple configurations as this is it practical to obtain an exact solution. Other more complex, and usually more interesting, configurations of guides will require approximations in order that the problem may be solved. These approximations will consequently be developed in this simple case and checked against the exact analysis to reveal the conditions under which the approximations are good approximations.

The type of approximation to be made is one which will be most useful in this and the following three chapters. It is based on the physical reasoning that the field distributions will not be changed greatly by the presence of a finite rather than an infinite conductivity. The major correction to the ideal analysis is the power loss due to current flow in the imperfect conductor, and this is calculated by using the currents and fields derived for ideal conductors. Other corrections may also be set down, such as a change in the value of wave impedance, and the addition

of an electric field component in the direction of current flow required by the imperfect conductor. It will turn out that for most practical cases it will be necessary only to calculate the power loss, but the other corrections will be studied here for purposes of enlightenment.

If the conducting plane which is to guide the wave has a finite conductivity, it must require some component of electric field in the direction of propagation to produce the current flow. This is given at once by the impedance of a plane conductor of infinite depth (Art. 6.07)

$$E_z = Z_s J_z = R_s J_z (1 + j)$$

With the value of J_z found in Eq. 8.02(3),

$$E_z = (1 + j) R_s H_y \qquad [1]$$

R_s is the skin effect surface resistivity defined in Art. 6.07.

The current flow in the plane of finite conductivity also produces a finite power loss determinable from the resistance component of the conductor impedance and the current flow in the plane.

$$P_L = \frac{J_z^2 R_s}{2} = \frac{H_y^2 R_s}{2} \quad \text{watts/meter}^2 \qquad [2]$$

Since a small component of E_z is given by (1), and since a component of energy flow into the plane accounts for ohmic losses, the wave may be considered as a plane wave with its wave front tipped slightly so that it is incident at a small but finite angle measured from the plane. The tangent of this angle is given approximately by the ratio of E_z to E_x. From (1) and Eq. 8.02(2), this ratio is

$$\frac{E_z}{E_x} = \frac{(1 + j) R_s}{\eta_1}$$

R_s is very small for reasonably good conductors at all radio frequencies, and η_1 is relatively large, so the above ratio is truly small. (For instance it is 3.8×10^{-5} for copper conductor, air dielectric, at 3000 mc; it is 3.8×10^{-6} at 30 mc.)

Since the wave front is tipped slightly, E_x is actually somewhat less than the total electric field, which is related to the magnetic field of the plane wave by η_1. This difference is extremely slight if the above ratio of R_s/η_1 is truly small. There must also be a phase velocity along the plane slightly different from the velocity of light, since the wave is incident at the small angle to the plane.

The above reasoning shows fairly clearly the type of approximations that may be made when conductors are imperfect. Since they will be

so useful in all subsequent analysis for guided waves along systems of finite conductivity, they are summarized below.

If conductors are imperfect, but reasonably good, it will usually be assumed that:

1. Transverse components of electric and magnetic field are practically the same as though conductors were perfect.

2. The power loss in the conductor will be found by assuming that for a given amount of a wave, current flow is practically the same as though the conductor were perfect. This current flow in the conductor of known resistance determines the power loss.

3. The component of electric field required to produce the above current is obtained from the current and the known internal impedance of the conductor. This is often completely negligible compared with other field components.

These are excellent approximations if:

1. Displacement currents in the conductor are negligible compared with conduction currents ($\sigma/\omega\epsilon \gg 1$). (Otherwise it is not correct to use merely R_s to determine E_z.)

2. The skin effect resistance of the conductor is extremely small compared with the intrinsic impedance of the dielectric ($R_s/\eta_1 \ll 1$). (Otherwise the relations between transverse components of electric and magnetic field will be appreciably disturbed.)

8.04 Exact Analysis for a Wave Guided by Imperfectly Conducting Plane

To solve exactly for the behavior of an electromagnetic wave guided by an imperfectly conducting plane as a check of the approximate results of the previous article, the proper solution of Maxwell's equations derived in the previous chapter corresponding to a wave that is incident upon a plane boundary could be chosen. However, it is desired to acquire the point of view that the waves are being guided along the boundary, and to develop a new technique for the study of guided waves. We shall again, therefore, start directly from Maxwell's equations. The curl relations in the dielectric (medium 1) and in the conductor (medium 2) are:

$$\text{Dielectric} \quad \nabla \times \bar{E} = -j\omega\mu_1 \bar{H} \qquad \nabla \times \bar{H} = j\omega\epsilon_1 \bar{E} \qquad [1]$$

$$\text{Conductor} \quad \nabla \times \bar{E} = -j\omega\mu_2 \bar{H} \qquad \nabla \times \bar{H} = j\omega\epsilon_c \bar{E} \qquad [2]$$

where

$$\epsilon_c = \epsilon_2\left(1 + \frac{\sigma_2}{j\omega\epsilon_2}\right) \qquad [3]$$

The general reasoning of the past article indicates that E_z should be included as well as the H_y and E_x components. Offhand we see no need for retaining other components, so we shall attempt to satisfy the conditions with these three components. If it is possible to satisfy all equations and boundary conditions, the number of components retained is sufficient. It will be seen that these three are truly sufficient for the present problem.

If there are no variations along the y coordinate, the component equations of (1) break into two independent sets, one of which relates the three components, E_x, H_y, and E_z (see Art. 4.26).

$$\begin{aligned} \frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} &= -j\omega\mu_1 H_y \\ -\frac{\partial H_y}{\partial z} &= j\omega\epsilon_1 E_x \\ \frac{\partial H_y}{\partial x} &= j\omega\epsilon_1 E_z \end{aligned} \qquad [4]$$

In Art. 7.02 it was shown that the curl equations must combine to give the wave equation in $\bar{E}$ or $\bar{H}$. In rectangular coordinates the wave equation may be written in terms of any of the components. For H_y,

$$\nabla^2 H_y = \mu_1\epsilon_1 \frac{\partial^2 H_y}{\partial t^2} = -\omega^2\mu_1\epsilon_1 H_y$$

For no variations with y,

$$\begin{aligned} \frac{\partial^2 H_y}{\partial x^2} + \frac{\partial^2 H_y}{\partial z^2} &= -k_1^2 H_y \\ k_1^2 = \omega^2\mu_1\epsilon_1 &= \frac{\omega^2}{v_1^2} \end{aligned} \qquad [5]$$

The technique to be employed in this and other guided wave studies is to assume at the beginning a function describing propagation in the direction of the guide, at the same time retaining the assumption of sinusoidal time variations. That is, it will be attempted to satisfy the differential equations and the boundary conditions with the function describing propagation in the z direction as $e^{j\omega t - \gamma z}$. γ is the propagation constant and will have in general real and imaginary parts, representing attenuation and propagation with a real velocity, respectively.

This function, substituted in the wave equation (5), leaves only

$$\frac{\partial^2 H_y}{\partial x^2} = -(\gamma^2 + k_1^2)H_y$$

Define

$$K_1^2 = -(\gamma^2 + k_1^2)$$

The solution to the above equation is in terms of exponentials in x.

$$H_y = e^{(j\omega t - \gamma z)}(C_1 e^{-K_1 x} + C_2 e^{K_1 x})$$

In the dielectric, the positive exponential must be absent if K has a positive real part; otherwise the fields would become infinite at $x = \infty$.

Note that since (1) and (2) have the same form, all equations obtained for the dielectric may be applied to the conductor if μ_1 is replaced by μ_2, and ϵ_1 is replaced by ϵ_c. However, the negative exponential must be absent in the solution which applies to the conductor; otherwise fields would become infinite at $x = -\infty$.

$$H_{y_1} = C_1 e^{-K_1 x} e^{(j\omega t - \gamma_1 z)} \qquad H_{y_2} = C_2 e^{K_2 x} e^{(j\omega t - \gamma_2 z)} \tag{6}$$

Other components follow from (4).

$$E_{x_1} = \frac{\gamma_1 C_1}{j\omega\epsilon_1} e^{-K_1 x} e^{(j\omega t - \gamma_1 z)} \qquad E_{x_2} = \frac{\gamma_2 C_2}{j\omega\epsilon_c} e^{K_2 x} e^{(j\omega t - \gamma_2 z)} \tag{7}$$

$$E_{z_1} = -\frac{K_1 C_1}{j\omega\epsilon_1} e^{-K_1 x} e^{(j\omega t - \gamma_1 z)} \qquad E_{z_2} = \frac{K_2 C_2}{j\omega\epsilon_c} e^{K_2 x} e^{(j\omega t - \gamma_2 z)} \tag{8}$$

The boundary conditions require that tangential electric and magnetic field components be continuous across the boundary (Arts. 4.22, 4.23).

$$\left.\begin{aligned} H_{y_1} &= H_{y_2} \\ E_{z_1} &= E_{z_2} \end{aligned}\right\} \quad x = 0$$

The continuity of H_y yields

$$C_1 e^{(j\omega t - \gamma_1 z)} = C_2 e^{(j\omega t - \gamma_2 z)}$$

This continuity must exist for all values of z; it can exist only if $\gamma_1 = \gamma_2$. Let this common propagation constant be denoted γ from now on. It also follows from the above relation that

$$C_2 = C_1 \tag{9}$$

The continuity of E_z, from (8), now yields

$$-\frac{K_1}{\epsilon_1} = \frac{K_2}{\epsilon_c} \tag{10}$$

From the definitions of K_1 and K_2,

$$K_1 = \sqrt{-(\gamma^2 + \omega^2\mu_1\epsilon_1)} \qquad K_2 = \sqrt{-(\gamma^2 + \omega^2\mu_2\epsilon_c)} \tag{11}$$

and (10), the propagation constant may be determined. Summarizing:

$$\gamma_1 = \gamma_2 = \gamma = \frac{\omega}{v_1}\sqrt{\frac{1 - \dfrac{\mu_2\epsilon_1}{\mu_1\epsilon_c}}{\left(\dfrac{\epsilon_1}{\epsilon_c}\right)^2 - 1}} \tag{12}$$

$$K_1 = \frac{\omega}{v_1}\frac{\epsilon_1}{\epsilon_c}\sqrt{\frac{\dfrac{\mu_2\epsilon_c}{\mu_1\epsilon_1} - 1}{\left(\dfrac{\epsilon_1}{\epsilon_c}\right)^2 - 1}} \tag{13}$$

$$K_2 = -\frac{\epsilon_c}{\epsilon_1}K_1 \tag{14}$$

8.05 Waves Along a Reasonably Good Conducting Plane

The exact expressions for propagation constant and field components for the wave guided by a conducting plane of any characteristics have been obtained, but they are not particularly meaningful in their complete form. In order to compare them with the approximate results of Art. 8.03, it is necessary to make the very practical and excellent approximation that displacement currents in the conductor are negligibly small compared with conduction currents. Results are then limited to reasonably good conductors. That is,

$$\frac{\sigma_2}{\omega\epsilon_2} \gg 1$$

$$\epsilon_c = \epsilon_2\left(1 + \frac{\sigma_2}{j\omega\epsilon_2}\right) \cong \frac{\sigma_2}{j\omega}$$

With these assumptions, it is consistent to assume also that the quantity $\sigma_2/\omega\epsilon_1$ is much greater than unity. The propagation constant and the constants K_1 and K_2 of Eqs. 8.04(12)–(14) accordingly reduce to

$$\gamma = \frac{k_1 R_s^2}{\eta_1^2} + jk_1\left[1 - \frac{2R_s^4\mu_1^2}{\eta_1^4\mu_2^2}\right] = \alpha + j\beta \tag{1}$$

$$K_1 = (1 - j)k_1\frac{R_s}{\eta_1} \tag{2}$$

$$K_2 = \frac{(1 + j)}{\delta_2} \tag{3}$$

where $k_1 = \omega\sqrt{\mu_1\epsilon_1}$, $\eta_1 = \sqrt{\mu_1/\epsilon_1}$, $\delta_2 = \dfrac{1}{\sqrt{\pi f \mu_2 \sigma_2}}$, $R_s = \sqrt{\pi f \mu_2/\sigma_2}$.

The expression for propagation constant, (1), shows that there is a small attenuation, α. This is small since $(R_s/\eta_1)^2$ is, for good conductors, an extremely small quantity. (For copper conductor, air dielectric $(R_s/\eta_1)^2$ is 0.48×10^{-12} at 1 mc, or 480×10^{-12} at 1000 mc.) There is also, as predicted in Art. 8.03, a phase velocity in the z direction slightly greater than the velocity of light.

$$v_p = \frac{\omega}{\beta} \cong v_1\left[1 + \frac{2R_s^4\mu_1^2}{\eta_1^4\mu_2^2}\right] \tag{4}$$

The difference between this velocity and the velocity of light is seen to be extremely small if R_s/η_1 is small.

Equations 8.04(6) and (7) give the value of wave impedance. This is

$$\frac{E_{x_1}}{H_y} = \frac{\gamma}{j\omega\epsilon_1} \cong \eta_1\left[1 - \frac{2R_s^4\mu_1^2}{\eta_1^4\mu_2^2} - j\frac{R_s^2}{\eta_1^2}\right] \tag{5}$$

For the condition of $R_s/\eta_1 \ll 1$, the difference between this ratio and that for the wave guided by the perfect conductor (ratio of $E_{x_1}/H_{y_1} = \eta_1$) is negligibly small.

From Eq. 8.04(8) and (6) E_z in the dielectric is given by

$$\frac{E_{z_1}}{H_{y_1}} = \frac{-K_1}{j\omega\epsilon_1} \cong (1 + j)R_s \tag{6}$$

This is exactly the same result as in Eq. 8.03(1).

Finally, the average power dissipated in the conductor per unit area may be found from the average value of the component of the Poynting vector normal to the plane.

$$P_x = H_y E_z\big|_{x=0}$$

$$\text{Time av } (P_x) = \tfrac{1}{2}(H_y^2 R_s) \quad \text{watts/meter}^2 \tag{7}$$

This is the same result as in Eq. 8.03(2).

It is found from the exact analysis that the conclusions listed at the end of Art. 8.03 are justified, and that the criteria determining the excellence of the approximations are that displacement currents in the conductor are negligible compared with conduction currents, and that the quantity $(R_s/\eta_1)^2$ be small compared with unity. The ratio R_s/η_1 may be expressed in at least two equivalent forms.

$$\frac{R_s}{\eta_1} = \frac{\pi\mu_2}{\mu_1}\frac{\delta}{\lambda_1} = \sqrt{\frac{\mu_2}{2\mu_1}\frac{\omega\epsilon_1}{\sigma_2}}$$

Before closing this discussion, there are several interesting characteristics evident from the previous results which are worth noting. E_z and E_x are 45° out of time phase so that the resultant electric vector will lean sometimes forward and sometimes slightly back of the normal. Its terminus describes an ellipse.

All quantities in the dielectric contain the factor

$$e^{-K_1 x} e^{(j\omega t - \gamma z)}$$

From (1) and (2) this may be written (neglecting $2R_s^4\mu_1^2/\eta_1^4\mu_2^2$ compared with unity)

$$e^{j\omega t} e^{-\left(\frac{k_1 R_s}{\eta_1}x + \frac{k_1 R_s^2}{\eta_1^2}z\right)} e^{-jk_1\left(z - \frac{R_s}{\eta_1}x\right)}$$

To find the planes of constant phase, it is necessary only to set

$$z - \frac{R_s x}{\eta_1} = \text{Constant}$$

These planes of constant phase then make an angle ϕ_1 with the plane,

$$\tan\phi_1 = \frac{\eta_1}{R_s} \qquad [8]$$

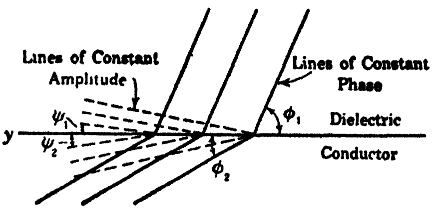


Fig. 8.05. Wave guided by an imperfect conductor.

The tilt of the planes is forward, as shown by the solid lines of Fig. 8.05. Actually this is much exaggerated, since it was found that R_s/η_1 is a very small quantity. ϕ_1 then differs from 90° by only a very small angle.

The planes of constant amplitude are obtained by setting

$$\frac{k_1 R_s x}{\eta_1} + \frac{k_1 R_s^2 z}{\eta_1^2} = \text{Constant}$$

These make an angle ψ_1 with the conductor, where

$$\tan\psi_1 = \frac{R_s^2\eta_1}{\eta_1^2 R_s} = \frac{R_s}{\eta_1} \qquad [9]$$

If R_s/η_1 is a small quantity, the planes of constant amplitude tilt backward and are very nearly parallel with the conductor, as shown by the dotted lines of Fig. 8.05. Similarly the planes of constant phase and amplitude in the conductor have angles ϕ_2 and ψ_2 respectively,

$$\tan\phi_2 = k_1\delta_2 = \frac{2\pi\delta_2}{\lambda_1} \qquad \tan\psi_2 = \frac{R_s^2 k_1\delta_2}{\eta_1^2} \qquad [10]$$

These lines are also sketched in Fig. 8.05. Both quantities are very small, but $\tan \psi_2 \ll \tan \phi_2$. Both planes of constant phase and constant amplitude in the conductor are then nearly parallel to the conductor surface, but the angles of the constant amplitude planes are even smaller than those of the constant phase planes. The difference between the planes of constant phase and planes of constant amplitude may be described in terms of a complex direction of propagation, the real part of which is normal to the planes of constant phase, and the imaginary part normal to the planes of constant amplitude. This concept is often used in analyses of waves in imperfect conductors.[1]

8.06 Transmission Line Type Wave between Parallel Planes

A second very simple case for which many important conclusions may be obtained without difficult mathematics is that of a wave guiding system consisting of a dielectric region between two parallel conducting planes of large extent (Fig. 8.06). If the planes may be considered as perfectly conducting, they impose the boundary condition that all electric fields tangential to the two conductors must be zero. It is apparent that this requirement is again met by a portion of a uniform plane wave, as in the case of the single perfectly conducting plane, Art. 8.02. So, a uniform plane wave with E_x only, H_y only, should propagate between the conducting planes in the z direction with a phase velocity equal to the velocity of light. The electric field passing normally between the plates corresponds to equal and opposite charge densities on the two plates at a given z plane; the uniform transverse magnetic field corresponds to equal and opposite currents flowing in the two plates in the z direction. This wave is then identified as the usual *transmission line type* of wave that was considered in the classical transmission line theory presented in Chapter 1. It is often called the *principal* wave.

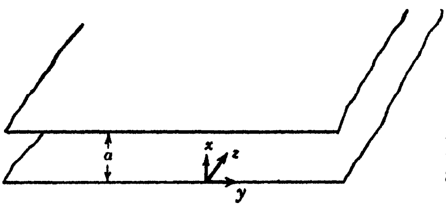


Fig. 8.06.

If the conductors are not perfect but actually have finite conductivity, it would seem safe now to make approximations so long as the conductors are good enough so that the requirements of Art. 8.03 are met. (That is, $\sigma_2/\omega\epsilon_2 \gg 1$ and $R_s/\eta_1 \ll 1$.) Then field components and currents are practically the same as though conductors were perfect. The major correction is the power loss in the conductors. This may be obtained by assuming that the same current as in the ideal case flows

[1] See for example Stratton, "Electromagnetic Theory," McGraw-Hill, 1941.

through the conductors of known conductivity. Total loss per unit length for the two planes is twice that for the single plane as calculated in Art. 8.03. For a width b of the transmission line,

$$W_L = b|H_y|^2 R_s \quad \text{watts/meter} \tag{1}$$

The average power transferred through a cross section normal to the planes is given by the average value of the axial component of Poynting vector, multiplied by the area of cross section.

$$W_T = \text{Av}[\bar{E} \times \bar{H}]_z ab = \frac{E_x H_y}{2} ab = \frac{H_y^2 ab}{2} \eta_1 \quad \text{watts} \tag{2}$$

Attenuation constant may be obtained from these two quantities as in Eq. 1.20(5)

$$\alpha = \frac{W_L}{2W_T} = \frac{R_s}{\eta_1 a} \quad \text{nepers/meter} \tag{3}$$

The ratio of the voltage between plates to the axial current is obtainable from the fields. Thus if we wish to find the voltage of the top plate with respect to the bottom,

$$V = -\int_0^a E_x\, dx = -aE_x$$

The current in the positive z direction in the upper plane, obtained from the $\bar{n} \times \bar{H}$ rule, is

$$I_z = bJ_z = -bH_y$$

So the ratio,

$$\frac{V}{I_z} = \frac{a}{b}\frac{E_x}{H_y} = \frac{a}{b}\eta_1 \tag{4}$$

If this wave is identical with that considered in conventional transmission line theory, these results should be obtainable by the theory of Chapter 1. At radio frequencies the transmission line formulas Eq. 1.17(6) reduce to

$$Z_0 \cong \sqrt{\frac{L}{C}} \cong \frac{1}{v_1 C}; \quad \alpha \cong \frac{R}{2Z_0}$$

where

C = capacitance per unit length.
R = resistance per unit length.
v_1 = velocity of light in the dielectric.
Z_0 = characteristic impedance, which gives the ratio of V/I_z for an infinite line.

For the parallel planes, $C = \epsilon_1 b/a$ farads per meter and $v_1 = 1/\sqrt{\mu_1\epsilon_1}$

$$Z_0 = \frac{1}{v_1 C} = \frac{a\sqrt{\mu_1\epsilon_1}}{\epsilon_1 b} = \frac{\eta_1 a}{b} \qquad [5]$$

This is identical with the ratio of V/I_z calculated from the wave solution in (4). The resistance per unit length of the line is

$$R = \frac{2R_s}{b}$$

so

$$\alpha = \frac{R}{2Z_0} = \frac{R_s}{a\eta_1} \qquad [6]$$

Then the attenuation constant also agrees with the results of the wave analysis. This is of course as it should be, since the two analyses are really equivalent for the transmission line wave under the approximations made here. We shall show later that this type of principal wave is in general one that can be accurately dealt with by the conventional transmission line operations, which has been convincingly proved by the fact that the same answer is obtained in that way as is gotten by the application of Maxwell's equations.

8.07 Higher Order Waves between Planes

The simple transmission line type of wave found in the previous article was a uniform plane wave propagating exactly parallel to the perfectly conducting planes. This is not the only wave that can propagate between conducting planes. If the expressions of Art. 7.11 for plane waves incident upon a conducting plane at some general angle θ are examined, it is noted that in this case there are other mathematical planes, removed from the conducting plane by $n_1\lambda_1/2 \cos\theta$ (θ is measured from the normal to the plane), for which electric field components tangential to the plane have become zero. Thus, for such a wave, two parallel perfectly conducting planes might be placed with just this separation and the required boundary conditions would be satisfied, This would correspond to the interference pattern caused by waves reflected at an angle θ, first, say, from the bottom conducting plane, then from the upper conducting plane, etc. (Fig. 8.07). Stating this in another way, for any given set of planes with arbitrary fixed spacing, there should be some frequencies and some angles of reflection for which the boundary conditions could be satisfied by a wave having a component of propagation in the z direction. There are any number of such waves,

corresponding to different values of n above. All have a definite pattern, which is not uniform in planes transverse to the conductors, so they may be thought of as *higher order* waves. They are sometimes called *complementary* waves as compared to the principal, or ordinary transmission line wave. Before looking at these waves further from the point of view of reflections, we shall go to the differential equations and determine what wave behavior may be expected for such waves.

It is again assumed that there are no variations in the y direction. The curl equations then are found to divide into two independent sets. One contains E_z, E_x, H_y only; the other contains H_z, H_x, E_y only. These represent two types of waves that may go on independently if

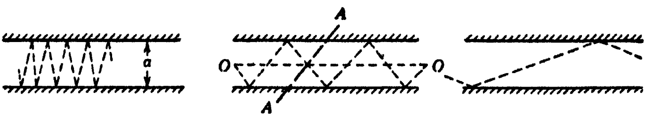


FIG. 8.07. Waves guided by two parallel conducting planes.

conductors are perfect, since there are no equations relating the two sets. The first set contains an $\bar{E}$ component but no $\bar{H}$ component in the direction of propagation; it may be called an E wave. Since it contains transverse components of magnetic field only, no axial component, it is also a *transverse magnetic* (TM) wave. Similarly, the second wave may be called an H wave or a *transverse electric* (TE) wave.

The two sets of component equations from the curl expressions, with $e^{(j\omega t-\gamma z)}$ substituted are

$$\begin{aligned} \gamma H_y &= j\omega\epsilon_1 E_x \\ \frac{\partial H_y}{\partial x} &= j\omega\epsilon_1 E_z \\ -\gamma E_x - \frac{\partial E_z}{\partial x} &= -j\omega\mu_1 H_y \end{aligned} \qquad [1]$$

$$\begin{aligned} \gamma E_y &= -j\omega\mu_1 H_x \\ \frac{\partial E_y}{\partial x} &= -j\omega\mu_1 H_z \\ -\gamma H_x - \frac{\partial H_z}{\partial x} &= j\omega\epsilon_1 E_y \end{aligned} \qquad [2]$$

The wave equations in terms of E_z and E_y:

$$\nabla^2 E_z = -k_1^2 E_z \text{ reduces to } \frac{\partial^2 E_z}{\partial x^2} = -(\gamma^2 + k_1^2)E_z$$

$$\nabla^2 E_y = -k_1^2 E_y \text{ reduces to } \frac{\partial^2 E_y}{\partial x^2} = -(\gamma^2 + k_1^2)E_y$$

$$k_1^2 = \omega^2 \mu_1 \epsilon_1$$

The solutions of these equations may be in sinusoids or exponentials in x. The requirement of zero E_z and E_y at the two perfectly conducting plates requires sinusoids for matching boundary conditions to obtain the repetition of zeros. The cosine term may also be eliminated if the bottom plate is taken as $x = 0$.

$$E_z = A \sin k_c x e^{(j\omega t - \gamma z)} \qquad [3]$$

$$E_y = B \sin k_c x e^{(j\omega t - \gamma z)} \qquad [4]$$

where we have defined

$$k_c^2 = k_1^2 + \gamma^2 \qquad [5]$$

Since $E_y = 0$ and $E_z = 0$ at $x = a$ as well as at $x = 0$, k_c is fixed such that there is a half period of the sine wave or a multiple thereof between the planes for either of the wave types,

$$k_c = \frac{n\pi}{a} \qquad [6]$$

n is any integer.

Equation (5) may be solved for γ:

$$\gamma = \sqrt{k_c^2 - k_1^2} \qquad [7]$$

Since k_c is a purely real number by (6), there is, for any given dimension a and any given integer n, some frequency at which γ is zero. This frequency may be called the cut-off frequency, f_c, corresponding to the dimension between plates and the particular wave order of interest. This is the frequency for which k_1 is exactly equal to k_c. The form is the same for both TM and TE waves.

$$f_c = \frac{v_1 k_c}{2\pi} = \frac{n v_1}{2a} \qquad [8]$$

This corresponds to a wavelength, measured in the dielectric if unbounded,

$$\lambda_c = \frac{v_1}{f_c} = \frac{2a}{n} \qquad [9]$$

The propagation constant may be written in terms of this cut-off frequency by substituting in (7).

$$\gamma = k_1\sqrt{(f_c/f)^2 - 1} \tag{10}$$

For frequencies less than the cut-off frequency $(f_c/f) > 1$, γ is a real number so that the propagation function $e^{-\gamma z}$ represents only attenuation. However, if frequency is greater than the cut-off value given by (8), $(f_c/f) < 1$, γ is a purely imaginary quantity,

$$\gamma = j\beta = jk_1\sqrt{1 - (f_c/f)^2} \tag{11}$$

The propagation function then corresponds to a propagation in the z direction with a real phase velocity and no attenuation.

$$v_p = \frac{\omega}{\beta} = \frac{v_1}{\sqrt{1-(f_c/f)^2}} \tag{12}$$

At the cut-off frequency, this velocity is infinite, but it decreases with frequency, and at very high frequencies it approaches the velocity of light in the medium. The phase velocity is always greater than the velocity of light in free space, just as was found in most past cases of waves traveling at an angle with respect to the direction in which velocity is measured. The group velocity, Eq. 1.25(4),

$$v_g = \frac{d\omega}{d\beta} = \frac{v_1^2}{v_p} = v_1\sqrt{1 - (f_c/f)^2} \tag{13}$$

Group velocity is zero at the cut-off frequency and increases with frequency, approaching the velocity of light and coinciding in magnitude with the phase velocity at very high frequencies.

Now that these propagation characteristics are definitely established, we may go back to the picture of waves reflected from the conducting planes at an angle, showing the consistency of results from the two points of view. The distance between planes of zero tangential electric field is, from the relations of Art. 7.11,

$$a = \frac{n\lambda_1}{2\cos\theta}$$

or

$$\cos\theta = \frac{n\lambda_1}{2a} = \frac{nv_1}{2af}$$

By comparing with (8), this is

$$\cos\theta = \frac{f_c}{f} \quad \text{and} \quad \sin\theta = \sqrt{1 - (f_c/f)^2} \tag{14}$$

So at a frequency equal to cut-off, the wave is passing normally between the planes and there is no component of energy flow in the direction parallel to the planes. At a frequency greater than f_c, θ is finite and the wave takes a path as in Figs. 8.07*a*, *b*, or *c*. There is now an energy flow in the z direction with velocity

$$\begin{aligned} v_g &= v_1 \sin\theta \\ &= v_1\sqrt{1 - (f_c/f)^2} \end{aligned}$$

The fictitious point of intersection of a plane of constant phase, as AA, and the z direction, OO, moves with velocity

$$v_p = \frac{v_1}{\sin\theta} = \frac{v_1}{\sqrt{1 - (f_c/f)^2}}$$

These agree exactly with the expressions for phase and group velocity determined previously in (12) and (13).

At very high frequencies, $\cos\theta$ is small, corresponding to propagation at flat angles or almost exactly in the direction of the guide. The wave fronts are then nearly normal to the direction of the guide.

Notice that the TM wave corresponds to a plane wave incident at the angle θ, polarized in the plane of incidence. A TE wave corresponds to one polarized normal to the plane of incidence. The change in angle θ then shows that most of the electric field for the TM wave would be in the z direction for frequencies near cut-off ($\cos\theta$ near unity) but in the transverse direction for very high frequencies. At any angle θ, the ratio of transverse E to H should be

$$\frac{E_T}{H_T} = \frac{\eta_1 H_y \sin\theta}{H_y} = \eta_1\sqrt{1 - (f_c/f)^2} \qquad [15]$$

For the TE wave, or wave polarized normal to the plane of incidence, magnetic field is entirely in the z direction at cut-off and is nearly all in the transverse direction at very high frequencies. At any angle θ,

$$\frac{E_T}{H_T} = \frac{E_y}{\dfrac{E_y}{\eta_1}\sin\theta} = \frac{\eta_1}{\sqrt{1 - (f_c/f)^2}} \qquad [16]$$

Exactly similar expressions could be obtained from the differential equation (2) with values of γ substituted from (11).

Below cut-off it is impossible to satisfy (14). That is to say that no angle of incidence will satisfy the boundary conditions and allow a plane

wave to exist without attenuation. The concept of reflected waves then requires special interpretation to be useful in studying the higher order waves below cut-off. We shall not bother with that here.

GENERAL ANALYSES OF GUIDED WAVES

8.08 Waves Guided by Uniform Systems

With the techniques and general feeling for guided wave behavior developed in the study of the preceding simple cases, we can now proceed more easily toward the analysis of a class of guides of more interest to engineers: coaxial lines, hollow pipe guides, parallel wire lines, etc. However, each of these is a special case of the general problem of the guiding of energy in a given direction by a uniform system of conductors; uniform, that is, in its geometrical configuration in the direction of energy guiding. Before taking up specific examples, we shall consequently derive certain basic relations and notions that apply to all uniform lines or guides.

To be completely general, the curl relations of Maxwell's equations with all components and all derivatives retained should be used to describe the electromagnetic behavior in the dielectric of the guiding system. The direction of propagation will be taken as the z direction, and it will be assumed that propagation behavior in this direction may be described by the function $e^{(j\omega t-\gamma z)}$. Any system of coordinates may be used in the plane normal to direction of propagation; because rectangular coordinates are used, let no one think that the discussion is limited to rectangular shapes of conductors.

The curl equations with the assumed functions $e^{(j\omega t-\gamma z)}$ are written below. The subscript 1 is used to signify quantities in the dielectric of the system.

$$\nabla \times \bar{E} = -j\omega\mu_1\bar{H} \qquad\qquad \nabla \times \bar{H} = j\omega\epsilon_1\bar{E}$$

$$\frac{\partial E_z}{\partial y} + \gamma E_y = -j\omega\mu_1 H_x \quad [1] \qquad\qquad \frac{\partial H_z}{\partial y} + \gamma H_y = j\omega\epsilon_1 E_x \quad [4]$$

$$-\gamma E_x - \frac{\partial E_z}{\partial x} = -j\omega\mu_1 H_y \quad [2] \qquad\qquad -\gamma H_x - \frac{\partial H_z}{\partial x} = j\omega\epsilon_1 E_y \quad [5]$$

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = -j\omega\mu_1 H_z \quad [3] \qquad\qquad \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = j\omega\epsilon_1 E_z \quad [6]$$

It must be remembered in all analysis to follow, that these coefficients E_z, H_x, E_y, etc., are functions of x and y only, by our agreement to take care of the z and time functions in the assumed $e^{(j\omega t-\gamma z)}$.

From the above equations, it is possible to solve for E_x, E_y, H_x, or H_y in terms of E_z and H_z.

$$H_x = \frac{1}{\gamma^2 + k_1^2}\left[j\omega\epsilon_1 \frac{\partial E_z}{\partial y} - \gamma \frac{\partial H_z}{\partial x}\right] \tag{7}$$

$$H_y = -\frac{1}{\gamma^2 + k_1^2}\left[j\omega\epsilon_1 \frac{\partial E_z}{\partial x} + \gamma \frac{\partial H_z}{\partial y}\right] \tag{8}$$

$$E_x = -\frac{1}{\gamma^2 + k_1^2}\left[\gamma \frac{\partial E_z}{\partial x} + j\omega\mu_1 \frac{\partial H_z}{\partial y}\right] \tag{9}$$

$$E_y = \frac{1}{\gamma^2 + k_1^2}\left[-\gamma \frac{\partial E_z}{\partial y} + j\omega\mu_1 \frac{\partial H_z}{\partial x}\right] \tag{10}$$

where $k_1^2 = \omega^2\mu_1\epsilon_1$.

If the dielectric has finite conductivity, σ_1, it is merely necessary to substitute $\epsilon_1\left(1 + \frac{\sigma_1}{j\omega\epsilon_1}\right)$ for ϵ_1 in the above expressions (Art. 7.15).

All waves propagating in the positive z direction according to the factor $e^{(j\omega t - \gamma z)}$ must have components related by these equations, since nothing has been assumed but this factor and Maxwell's equations. [For a wave traveling in the negative z direction, substitute $-\gamma$ for γ in (1)–(6) or (7)–(10).] The total electric and magnetic intensities in the charge-free regions between the conducting boundaries must also satisfy the wave equation (Art. 7.02).

$$\nabla^2\bar{E} = -k_1^2\bar{E} \qquad \nabla^2\bar{H} = -k_1^2\bar{H}$$

The three-dimensional ∇^2 may be broken into two parts

$$\nabla^2\bar{E} = \nabla_{xy}^2\bar{E} + \frac{\partial^2\bar{E}}{\partial z^2}$$

The last term is the contribution to ∇^2 from derivatives in the axial direction. The first term is the two-dimensional Laplacian in the transverse plane, representing contributions to ∇^2 from derivatives in this plane. By the assumed propagation function, $e^{-\gamma z}$, in the axial direction,

$$\frac{\partial^2\bar{E}}{\partial z^2} = \gamma^2\bar{E}$$

The above wave equations may then be written

$$\nabla_{xy}^2\bar{E} = -(\gamma^2 + k_1^2)\bar{E} \tag{11}$$

$$\nabla_{xy}^2\bar{H} = -(\gamma^2 + k_1^2)\bar{H} \tag{12}$$

Equations (11) and (12) are the differential equations that must be satisfied in the dielectric region bounded by the conductors of the transmission lines or guides. The boundary conditions imposed on these differential equations follow from the configuration and the electrical properties of the conducting guides. Equations (1) to (6) or (7) to (10) then give the relations between any desired components in the wave.

It will be advantageous to be quite general in studying the types of waves that may propagate along the uniform guide, but the generalizations will be chosen carefully. We recognize, for instance, that any solution to the wave equation, and so the most general type of propagating wave imaginable, may be built up by superposing other simpler solutions of the wave equation. One example, in waves guided between planes, has already been encountered in which the more complex waves were built up from the simplest possible wave solutions: uniform plane waves propagating with the velocity of light in the dielectric. The trick was, of course, to add these waves in proper amounts and with proper directions of propagation. However, this method has limitations in usefulness because of the numbers of the different waves and different directions of propagation that must be retained if the reflections from the guiding geometry are at all complex. For the present study it will be desired to build up all wave types from basic plane waves defined as uniform in phase, though generally non-uniform in amplitude, over the cross-sectional plane, and propagating in the axial direction. The first division into types (which have already been met in the parallel plane guides) is as follows.

1. Waves that contain neither electric nor magnetic field in the direction of propagation. Since electric and magnetic field lines both lie entirely in the transverse plane, these may be called *transverse electromagnetic waves* (abbreviated *TEM*). They are the usual *transmission line waves* and are sometimes known as the *principal waves.*

2. Waves that contain electric field but no magnetic field in the direction of propagation. These are known as *E waves*, or *transverse magnetic waves* (*TM*).

3. Waves that contain magnetic field but no electric field in the direction of propagation. These are known as *H waves* or *transverse electric waves* (*TE*).

This is one way of dividing up possible wave types and, of course, it is often possible to have many waves of each type present at once along a given guiding system, although it will be found that certain types of guiding systems will not allow all the above types. Any one of the allowed waves may exist by itself if it is excited, and if conditions are favorable for its propagation. However, there may be more complex

propagating waves, and these can be considered as made up of proper amounts of the separate basic waves.

We shall now proceed to a study of these general types separately. The general analysis follows quite closely that given by Schelkunoff.[2]

8.09 Transverse Electromagnetic or Transmission Line Waves

The first of the basic wave types to be studied is that with neither electric nor magnetic field in the direction of propagation. This has been termed a transverse electromagnetic wave. In the simple case of propagation between perfectly conducting parallel planes, such a wave was identified exactly with the ordinary wave expected from transmission line theory. It will now be shown that this must be true for any general cross section of a uniform guiding line with perfect conductors along which this wave type may exist. (The types of guides on which it may not exist will be apparent once its characteristics are found.)

The general relations between wave components as expressed by Eq. 8.08(7) to Eq. 8.08(10) show that with E_z and H_z zero, then all other components must of necessity also be zero, unless $\gamma^2 + k_1^2$ is at the same time zero. Thus, a transverse electromagnetic wave must satisfy the condition

$$\gamma^2 + k_1^2 = 0$$

or

$$\gamma = \pm jk_1 = \pm j\frac{\omega}{v_1} = \pm j\omega\sqrt{\mu_1\epsilon_1} \qquad [1]$$

For a perfect dielectric, the propagation constant γ is thus a purely imaginary quantity, signifying that any completely transverse electromagnetic wave must propagate unattenuated, and with velocity v_1, the velocity of light in the dielectric bounded by the guide.

With (1) satisfied, the wave equations, as written in the form of Eqs. 8.08(11) and 8.08(12), reduce to

$$\nabla_{xy}^2\bar{E} = 0 \qquad \nabla_{xy}^2\bar{H} = 0 \qquad [2]$$

These are exactly the form of the two-dimensional Laplace's equation written for $\bar{E}$ and $\bar{H}$ in the transverse plane. Since E_z and H_z are zero, $\bar{E}$ and $\bar{H}$ lie entirely in the transverse plane. In Art. 3.02 it was found that electric and magnetic fields both satisfy Laplace's equation under

[2] Schelkunoff, "Transmission Theory of Plane Electromagnetic Waves," *Proc. I.R.E.*, 25, 1457 (November, 1937).

static conditions. Consequently it may be concluded that the field distribution in the transverse plane is exactly a static distribution, if it can be shown that boundary conditions to be applied to the differential equations (2) are the same as those for a static field distribution. The boundary condition requires that the electric field at the surface of the conductor have normal components only. The line integral of this electric field between the two conductors, evaluated in the transverse plane, may be used as a voltage or potential difference between the lines, corresponding to equal and opposite charges on the two conductors at that plane. This is valid because $\bar{E}$ does indeed satisfy Laplace's equation, and so may be thought of as the gradient of a scalar potential which here corresponds to voltage.

If electric field components are normal to the surface of perfect conductors, magnetic field components are entirely tangential. We can show this by noting from Eq. 8.08(1) and Eq. 8.08(4) that if E_z and H_z are zero

$$H_y = \frac{j\omega\epsilon_1}{\gamma} E_x = \frac{E_x}{\eta_1} \tag{3}$$

and

$$H_x = -\frac{\gamma}{j\omega\mu_1} E_y = -\frac{E_y}{\eta_1} \tag{4}$$

[The signs of (3) and (4) are for a positively traveling wave; for a negatively traveling wave they are opposite.] Study shows that (3) and (4) are conditions which require that electric and magnetic field be everywhere normal to each other. In particular, magnetic field must be tangential to the conducting surfaces since electric field is normal to them. The magnetic field pattern in the transverse plane then corresponds exactly to that arising from static currents flowing entirely on the surfaces of the perfect conductors.

The above characteristics show that a transverse electromagnetic wave may be guided by two or more conductors, or outside a single conductor, but not inside a closed conducting region, since it can have only the distributions of the corresponding two-dimensional static problem.

The above conditions also show that it is possible and completely correct to analyze the behavior of this wave from the conventional transmission line point of view, using distributed capacities and inductances calculated from D-C conditions. To complete this identity of viewpoints, let us arrive at the usual equations written in terms of voltages and currents, starting from the known field components.

In order that a definite example may be referred to, consider a line consisting of two conductors A and B of any general shape, Fig. 8.09. We shall, for the demonstration, be quite general regarding time and z functions, merely requiring that E_z and H_z be zero. The voltage between the two lines may be found by integrating electric field over any path between lines, such as that shown, 1-0-2. It will have the same value no matter what path is chosen, since $\bar{E}$ does satisfy Laplace's equation in the transverse plane and so may be considered as the gradient of a scalar potential in so far as variations in the transverse plane are concerned.

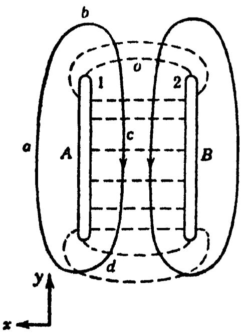


FIG. 8.09. Two-conductor transmission line with integration paths.

$$V = -\int_1^2 \bar{E} \cdot d\bar{l} = -\int_1^2 (E_x\,dx + E_y\,dy)$$

Differentiate the above equation with respect to z.

$$\frac{\partial V}{\partial z} = -\int_1^2 \left(\frac{\partial E_x}{\partial z}\,dx + \frac{\partial E_y}{\partial z}\,dy \right)$$

But the curl relation,

$$\nabla \times \bar{E} = -\frac{\partial \bar{B}}{\partial t}$$

shows that if E_z is zero,

$$\frac{\partial E_y}{\partial z} = \frac{\partial B_x}{\partial t} \quad \text{and} \quad \frac{\partial E_x}{\partial z} = -\frac{\partial B_y}{\partial t}$$

By substituting these in the above equation,

$$\frac{\partial V}{\partial z} = -\frac{\partial}{\partial t}\int_1^2 (-B_y\,dx + B_x\,dy)$$

A study of Fig. 8.09 reveals that the quantity inside the integral is the magnetic flux flowing across the path 1-0-2, per unit length in the z direction. According to the usual definition of inductance, this may be written as the product of inductance L per unit length and the current I.

$$\frac{\partial V}{\partial z} = \frac{\partial}{\partial t}(LI) = -L\frac{\partial I}{\partial t} \qquad [5]$$

The above is one of the differential equations used as a starting point for conventional transmission line analysis (Art. 1.15). The other may

be developed by starting with current in line A as the integral of magnetic field about a path a-b-c-d-a. (There is no contribution from displacement current since there is no E_z.)

$$I = \oint \bar{H} \cdot d\bar{l} = \oint (H_x\, dx + H_y\, dy)$$

Differentiate with respect to z.

$$\frac{\partial I}{\partial z} = \oint \left(\frac{\partial H_x}{\partial z}\, dx + \frac{\partial H_y}{\partial z}\, dy \right)$$

From the curl equation,

$$\nabla \times \bar{H} = \frac{\partial \bar{D}}{\partial t}$$

it follows that if $H_z = 0$,

$$\frac{\partial H_y}{\partial z} = -\frac{\partial D_x}{\partial t} \quad \text{and} \quad \frac{\partial H_x}{\partial z} = \frac{\partial D_y}{\partial t}$$

Substituting,

$$\frac{\partial I}{\partial z} = -\frac{\partial}{\partial t} \oint (D_x\, dy - D_y\, dx)$$

Inspection of the figure shows that this must be the electric displacement flux per unit length of line crossing from one conductor to the other. Since it corresponds to the charge per unit length on the conductors, it may be written as the product of capacity per unit length and the voltage between lines.

$$\frac{\partial I}{\partial z} = -C \frac{\partial V}{\partial t} \tag{6}$$

Equations (5) and (6) are exactly the equations used as a beginning for transmission line analysis, neglecting losses. (Art. 1.15.) It is seen that they may be derived exactly from Maxwell's equations provided the conductors are perfect. So, in this very important case of guiding of electromagnetic energy (transmission lines with negligible imperfections in conductivity of conductors), the well-known method of analysis based upon low-frequency circuit notions gives the correct answer, since it is actually equivalent to an analysis starting from Maxwell's equations, this despite the use of *static* L's and C's for a problem certainly not static. It will be seen, however, that there are many other types of guides for which the rigorous wave equations will appear in

quite a different role. Instead of serving as a means of proving the L and C approach to be legitimate, for these guides the field approach will prove to be the only effective means of analysis, at least until such rigorous field analyses divulge a way of properly extending inductance and capacitance concepts to these more complex cases.

Problem 8.09. Demonstrate that although in a *TEM* wave $\bar{E}$ does satisfy Laplace's equation in the transverse plane and so may be considered as a gradient of a scalar in so far as variations in the transverse plane are concerned, $\bar{E}$ is *not* the gradient of a scalar when variations in all directions (x, y, and z) are included.

8.10 Transmission Line Waves Along Imperfect Lines

It has now been found that a wave with neither electric nor magnetic field components in the direction of propagation may propagate along a perfect transmission line with the velocity of light in the dielectric surrounding the line. The fields in the transverse plane have exactly a static field distribution. Moreover, the classical analysis for such a transmission line wave, made in terms of voltage and current along the line and the distributed inductance and capacitance calculated at DC, is equivalent to one made directly from Maxwell's equations. This conclusion might not have been expected, for if one had wished to be skeptical, it would have been easy to question the validity of the transmission line equations on at least two counts.

1. A voltage drop due to current flow through the distributed inductance of the line is calculated, but none is included because of mutual effects from any other part of the line; similarly, no mutual charging effects are considered.

2. Inductance and capacitance used in the equations are those calculated for DC. It might seem doubtful that such constants could be of any use for extremely high frequencies; certainly we found that it is not permissible to neglect frequency effects when considering lumped inductances and capacitances at the highest frequencies in circuit equations (Chapter 5).

The first objection is answered once it is found from the field equations that there are no axial field components in the wave, and consequently no mutual effects. The second objection is answered by the discovery that the field distribution for the wave in the transverse plane is actually one corresponding to the static field pattern for that configuration, no matter what the frequency may be. The necessary condition is that the propagation be with light velocity in the dielectric of the line, a condition the conventional approach to transmission lines is very happy to grant.

If the transmission line is not ideal, but has resistance and conductance

of finite amount, classical transmission line theory would have us take account of these by setting the voltage change along the line equal to a resistance plus an inductance drop, and the current change equal to a capacitance plus a conductance leakage current (Art. 1.24).

$$\frac{\partial V}{\partial z} = -(j\omega LI + RI)$$

$$\frac{\partial I}{\partial z} = -(j\omega CV + GV)$$

It is usually assumed that inductance and capacitance calculated on the basis of D-C distributions are still used in these equations. Although it is true that a contribution to inductance from the flux inside the conductors (the internal inductance of Chapter 6) may now be included, that part of the inductance arising from flux in the space between conductors is still calculated from the D-C distributions.

It will now be shown that such an analysis is equivalent to one made from Maxwell's equations if a line has uniform conductance but no resistance; it will also be shown that if resistance of the conductors is important, the two analyses cannot be exactly equivalent. However, we should not undermine our confidence in the usual transmission line expressions too quickly, for the error will be infinitesimal for efficient transmission lines.

If the transmission line has a dielectric with uniform conductivity σ_1, occupying all the space between conductors, previous field analyses can be corrected by replacing $j\omega\epsilon_1$ by $(\sigma_1 + j\omega\epsilon_1)$ in all results (Art. 7.15). However, this is exactly what is done in a conventional analysis, where $j\omega C$ for the ideal line is replaced by $(G + j\omega C)$ for the line with conductance. For a line with uniform dielectric, G has the same form as C, with conductivity in place of dielectric constant. For example, in a coaxial line

$$G = \frac{2\pi\sigma_1}{\ln\left(\dfrac{r_0}{r_i}\right)} \quad \text{mhos/meter}$$

$$C = \frac{2\pi\epsilon_1}{\ln\left(\dfrac{r_0}{r_i}\right)} \quad \text{farads/meter}$$

Or, in general,

$$(G + j\omega C) = (\sigma_1 + j\omega\epsilon_1) \times \text{Function of configuration}$$

It follows that the two analyses have then actually considered the effect of conductivity of the dielectric in the same manner.

To clear up a few questions that may remain from the above demonstration, note that the quantity $\gamma^2 + \omega^2\mu_1\epsilon_1$ appearing in the wave equations now becomes

$$\gamma^2 + \omega^2\mu_1\epsilon_1\left(1 + \frac{\sigma_1}{j\omega\epsilon_1}\right)$$

That is,

$$\nabla^2_{xy}\bar{E} = \left[\gamma^2 + \omega^2\mu_1\epsilon_1\left(1 + \frac{\sigma_1}{j\omega\epsilon_1}\right)\right]\bar{E}$$

This quantity may be zero as before, since γ will now have attenuation as well as a phase constant. That is,

$$\gamma = j\omega\sqrt{\mu_1\epsilon_1}\sqrt{1 + \frac{\sigma_1}{j\omega\epsilon_1}}$$

will reduce the above equation to Laplace's equation. Then the same conclusions found previously apply. There can be no axial components, E_z or H_z; the fields in the transverse plane satisfy Laplace's equation. This again checks the validity of the transmission line analysis which neglects mutuals and uses distributed constants calculated at DC. There is, to be sure, some attenuation of power because of the imperfect dielectric, but it will be correctly computed by the ordinary line equations.

If the current-carrying conductors of the transmission line have finite conductivity, one trouble is immediately apparent. There must be at least some small component of electric field in the direction of propagation to force the current through the conductors. By referring again to Eqs. 8.08(7) to 8.08(10), it is seen that with E_z finite, $\gamma^2 + k_1^2$ must then also be finite. The quantity on the right of the wave equation cannot then be exactly zero, but must be some small but finite amount.

$$\nabla^2_{xy}\bar{E} = \text{Finite quantity}$$

This indicates that the field distributions are disturbed from the Laplace distributions somewhat by the axial field required to produce current flow. It is then no longer correct to calculate values of capacitance and inductance from the static distributions.

Although the nature of an exact analysis from Maxwell's equations is

apparent, it is difficult to apply to practical lines. One must first obtain the wave solutions which apply inside the dielectric and those which apply inside the conductor, matching the two at the boundary. The difficulties with most geometrical configurations are obvious. Schelkunoff has carried through this attack for coaxial lines,[3] determining the extent of the approximations which must be made to reduce the problem to the classical analysis. Studies of more general configurations are made by the method of successive perturbations. That is, the first correction to the perfect conductor case is the required axial electric field, which may be estimated simply from the resistivity times the approximate current flow. An idea is thus obtained of E_z's distribution and magnitude and consequently of $\nabla^2 E_z$. A next approximation is then obtained for the distribution of E_x, H_y, etc., as well as γ. From the new H's thus computed, a new current is computed and the whole process is again repeated. From the results of such studies it becomes apparent that an exact analysis from Maxwell's equations is fortunately unnecessary for lines which are at all efficient for energy transfer. The difference in results between such an exact analysis and the usual classical analysis including distributed resistance is extremely small.[4]

The classical transmission line analysis for imperfectly conducting boundaries is similar to methods previously introduced in this book, in which the first correction arising from the resistance is applied, but the major field distributions are assumed essentially unchanged. When this type of approximation was used for a wave analysis in Art. 8.03, the two criteria for its use were

1. Displacement currents in the conductor negligible compared to conduction currents.

2. The intrinsic impedance of the dielectric much greater than the skin effect surface resistivity of the conductor.

These are also a measure of the excellence of the conventional transmission line analysis including distributed resistance. Stated in another way, such an analysis assumes that transverse electric field components *in the conductor* are negligible compared with the axial, and that axial electric field components *in the dielectric* are small compared with the transverse. These are equivalent to the above. Thus

$$\frac{\sigma_2}{\omega\epsilon_2} \gg 1$$

[3] S. A. Schelkunoff, "The Electromagnetic Theory of Coaxial Transmission Lines and Cylindrical Shields," *Bell System Tech. Journ.*, 13, 532 (October, 1934).

[4] J. R. Carson, "The Guided and Radiated Energy in Wire Transmission," *Journ. A.I.E.E.*, pp. 906–913, October, 1914.

and

$$\frac{R_{s_2}}{\eta_1} \ll 1$$

where

$$R_{s_2} = \sqrt{\frac{\pi f \mu_2}{\sigma_2}} \quad \text{and} \quad \eta_1 = \sqrt{\frac{\mu_1}{\epsilon_1}}$$

The subscript 1 denotes dielectric, and 2 the conductor. If these inequalities are not satisfied, one must then examine critically any results predicted by the usual transmission line equations.

8.11 Transverse Magnetic or *E* Waves

In the general conducting guide of uniform section, the first possibility, that of a wave with neither electric nor magnetic field in the direction of propagation, was found to be the familiar wave of conventional transmission lines. As the next possibility, let us consider what waves may exist with electric field but no magnetic field in the direction of propagation. Because of the presence of axial electric field, these have been called *E waves.*. Because magnetic field lies entirely in a plane transverse to direction of propagation, the *E* waves are also called *transverse magnetic* (*TM*) *waves.* Before studying physically what such waves represent, let us first determine the characteristics they must have to satisfy Maxwell's equations and the boundary conditions imposed by the conducting guide which will first be assumed perfectly conducting.

With the usual assumed propagation constant, $e^{(j\omega t - \gamma z)}$, the axial component of electric field which is now present must satisfy the wave equation in the form of Eq. 8.08(11).

$$\nabla_{xy}^2 E_z = -(\gamma^2 + k_1^2)E_z$$

where

$$k_1^2 = \omega^2 \mu_1 \epsilon_1 \qquad [1]$$

The quantity $(\gamma^2 + k_1^2)$ should be a constant for any given type of wave. Moreover, the value of this constant is determined by the frequency and the boundary conditions imposed by the conducting guide, as was found in the study of waves between parallel planes. Let us call this constant k_c^2.

$$\nabla_{xy}^2 E_z = -k_c^2 E_z$$

$$k_c^2 = \gamma^2 + k_1^2 \qquad [2]$$

The next step is to study the complete boundary conditions that must be applied to solutions of this equation in order to determine k_c.

Assuming that the conducting guides are perfect, E_z at least must be zero at the conducting boundaries of the guide. There is naturally a question as to whether this is all the boundary condition that is required, or whether some other condition must be imposed to insure that tangential electric field components in the transverse plane also disappear at the surface of the conductor. It will next be shown that the single requirement of $E_z = 0$ at the conductor is a sufficient boundary condition, and at the same time further information about the wave will be given. Let us first write all field components from the general equations of Art. 8.08 with H_z set equal to zero. (Relations are for a positively traveling wave; for a negatively traveling wave, change the sign of γ.)

$$H_x = \frac{j\omega\epsilon_1}{k_c^2}\frac{\partial E_z}{\partial y} \qquad E_x = -\frac{\gamma}{k_c^2}\frac{\partial E_z}{\partial x}$$
$$H_y = \frac{-j\omega\epsilon_1}{k_c^2}\frac{\partial E_z}{\partial x} \qquad E_y = -\frac{\gamma}{k_c^2}\frac{\partial E_z}{\partial y} \qquad [3]$$

From the above

$$E_x = \frac{\gamma}{j\omega\epsilon_1}H_y \qquad E_y = -\frac{\gamma}{j\omega\epsilon_1}H_x \qquad [4]$$

A study of these relations shows first that the component of electric field in the transverse plane is normal to the component of magnetic field in the transverse plane. It follows next that the transverse component of magnetic field (which is the only magnetic field component for a TM wave) is normal to the total component of electric field.

Recall next that magnetic field must be tangential to the perfectly conducting boundary at all points (see Art. 4.23). Since total electric field has been shown to be normal to the magnetic field, this total electric field is normal to the boundary. This is the only requirement for completely satisfying the boundary conditions imposed by the perfect conductor.

Thus the single condition of $E_z = 0$ along perfectly conducting boundaries is the complete boundary condition. The constant k_c is determined from this boundary condition, thus automatically fixing the propagation constant γ.

$$\gamma = \sqrt{k_c^2 - k_1^2} \qquad [5]$$

Problem 8.11. Show that lines of constant E_z in the transverse plane are coincident with magnetic field lines.

8.12 Propagation Characteristics of *TM* Waves in Ideal Guides

Several limitations may be found for k_c in Eq. 8.11(2) which will affect the type of propagation constant possible for the wave. For one thing, if we now limit ourselves to waves in which all points in the transverse plane are in phase, then the derivatives in the transverse plane, as expressed by ∇^2_{xy}, can bring in no imaginary or complex numbers. The constant k_c^2 must then be entirely real.

It can also be shown that k_c^2 must be a positive number if the wave is guided inside a system completely enclosed by conductors. For this purpose, use the divergence theorem (Art. 2.12) written in a form applicable to the two-dimensional case.

$$\int_{\text{c.s.}} (\nabla \cdot \bar{F})dS = \oint F_n \, dl$$

$\bar{F}$ is any vector which does not vary in the axial direction; the first integral is the integral over the area of the cross section; the second is the line integral about the boundary. All vector operations are confined to the transverse plane. Since E_z is not a function of z but merely the coefficient of $e^{(j\omega t - \gamma z)}$, $\bar{F}$ may be taken as the vector $(E_z \nabla E_z)$. The integral on the right then disappears if the region is completely enclosed by a conductor (as in hollow pipe wave guides) since $E_z = 0$ along the conducting boundary.

$$\int_{\text{c.s.}} \nabla \cdot (E_z \nabla E_z)dS = 0$$

The quantity inside the integral may be expanded (Art. 2.38):

$$\int_{\text{c.s.}} [(\nabla E_z)^2 + E_z \nabla^2 E_z]dS = 0$$

The value of $\nabla^2 E_z$ is supplied by Eq. 8.11(2). Then

$$\int_{\text{c.s.}} (\nabla E_z)^2 dS = k_c^2 \int_{\text{c.s.}} E_z^2 dS \qquad [1]$$

Since E_z^2 and ∇E_z^2 are always real and positive, k_c^2 must also be always real and positive.

From Eq. 8.11(5), the propagation constant is

$$\gamma = \sqrt{k_c^2 - k_1^2} \qquad [2]$$

For $k_c > k_1$, γ is a purely real quantity and so represents attenuation. For $k_c < k_1$, γ is a purely imaginary quantity and represents propagation with a real velocity and no attenuation. The condition under which

γ becomes zero, as the transition between the propagating and attenuating behavior is reached, may be called the cut-off condition for the guide. Thus cut-off is at $k_1 = k_c$ or

$$f_c = \frac{k_c v_1}{2\pi} = \frac{k_c}{2\pi\sqrt{\mu_1\epsilon_1}} \qquad [3]$$

$$\lambda_c = \frac{2\pi}{k_c} \qquad [4]$$

Equation (2) may now be written in terms of the cut-off frequency.

$$\gamma = \alpha = k_c\sqrt{1 - (f/f_c)^2} \quad f < f_c \qquad [5]$$

$$\gamma = j\beta = jk_1\sqrt{1 - (f_c/f)^2} \quad f > f_c \qquad [6]$$

The phase velocity corresponding to the real velocity of propagation above cut-off is

$$v_p = \frac{\omega}{\beta} = \frac{v_1}{\sqrt{1 - (f_c/f)^2}} \qquad [7]$$

The group velocity,

$$v_g = \frac{d\omega}{d\beta} = v_1\sqrt{1 - (f_c/f)^2} \qquad [8]$$

Phase velocity is infinite at cut-off frequency and is always greater than the velocity of light in the dielectric; group velocity is zero at cut-off and is always less than the velocity of light in the dielectric. As the frequency increases beyond cut-off, phase and group velocities both approach the velocity of light in the dielectric.

The above expressions for phase and group velocities are the general ones, having the same form for *TM* waves in any enclosed uniform guiding system, although of course f_c will be different for each shape and size of guide. This form was obtained in considering the special case of Art. 8.07 (higher order waves between parallel planes). We could not have predicted from that single example that the form would be found so general.

8.13 Characteristic Wave Impedance of *TM* Waves

In Chapter 7, the concept of a wave impedance based upon the ratio of the transverse component of electric field to the transverse component of magnetic field, was found to be extremely useful in problems of trans-

mission and reflection at discontinuities. For the transverse magnetic or E wave, it is evident from Eq. 8.11(4) that this ratio is a constant.

$$\frac{E_T}{H_T} = \frac{\sqrt{E_x^2 + E_y^2}}{\sqrt{H_x^2 + H_y^2}} = \frac{\gamma}{j\omega\epsilon_1}$$

This ratio may be considered as the characteristic wave impedance for the transverse magnetic or E wave.

$$Z_{TM} = \frac{\gamma}{j\omega\epsilon_1} \tag{1}$$

If the value of γ from Eq. 8.12(6) is substituted,

$$\begin{aligned} Z_{TM} &= \eta_1\sqrt{1 - (f_c/f)^2} \\ \eta_1 &= \sqrt{\mu_1/\epsilon_1} \end{aligned} \tag{2}$$

This impedance is imaginary (reactive) below the cut-off frequency, zero at cut-off, and purely real above cut-off, approaching the intrinsic impedance of the dielectric at infinite frequency. This is a familiar behavior for the characteristic impedance of filter sections and again emphasizes the cut-off properties of the wave. That is, since Z_{TM} is purely reactive below cut-off, the wave can produce no net energy flow down the guide. Above cut-off it is real, allowing energy propagation.

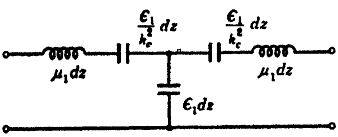


FIG. 8.13. Equivalent circuit for the transverse magnetic wave.

Problem 8.13. Show that the circuit of Fig. 8.13 may be used to represent the propagation characteristics of the transverse magnetic wave, if the characteristic wave impedance and propagation constant are written by analogy with transmission line results in terms of an impedance Z_1, and an admittance Y_1 per unit length.

$$Z_{TM} = \sqrt{Z_1/Y_1} \qquad \gamma = \sqrt{Z_1Y_1}$$

Note the similarity between this and the circuits of conventional filter sections, remembering of course that all constants in this circuit are in reality distributed constants.

8.14 Power Transfer in Wave Guides with *TM* Waves

The power transfer down the guide has been shown to be zero below cut-off, if the conductor of the guide is perfect. Above cut-off it may be obtained in terms of the field components by integrating the axial component of Poynting vector over the cross-sectional area.

Since it was shown that transverse components of electric and magnetic field are in phase and normal to each other, the axial component of the Poynting vector must be merely the product of these transverse field components.

$$\bar{P} = \bar{E} \times \bar{H}$$

Time average
$$P_z = \frac{|E_T||H_T|}{2}$$

E_T and H_T (transverse components) are related by the characteristic wave impedance, Z_{TM} so

$$P_z = \frac{Z_{TM}}{2} |H_T|^2 \tag{1}$$

But from Eq. 8.11(3)

$$|H_T|^2 = |H_x^2 + H_y^2| = \frac{\omega^2\epsilon_1^2}{k_c^4}\left[\left(\frac{\partial E_z}{\partial y}\right)^2 + \left(\frac{\partial E_z}{\partial x}\right)^2\right]$$

The factor in the brackets is merely the square of the two-dimensional gradient of E_z. By substituting,

$$H_T^2 = \frac{\omega^2\epsilon_1^2}{k_c^4} [\nabla E_z]^2 \tag{2}$$

Then the integral of P_z over the cross section is

$$W_T = \int_{\text{c.s.}} \frac{Z_{TM}}{2} |H_T|^2 \, dS = \frac{Z_{TM}\omega^2\epsilon_1^2}{2k_c^4} \int_{\text{c.s}} [\nabla E_z]^2 \, dS$$

By substitution from Eq. 8.12 (1),

$$W_T = \frac{Z_{TM}\omega^2\epsilon_1^2}{2k_c^2} \int_{\text{c.s.}} E_z^2 dS$$

$$= \frac{Z_{TM}}{2\eta_1^2}\left(\frac{f}{f_c}\right)^2 \int_{\text{c.s.}} E_z^2 dS \quad \text{watts} \tag{3}$$

Notice that this form of the integral seems to indicate that energy transfer increases with square of frequency at frequencies far enough above cut-off so that Z_{TM} is substantially constant. This is true if E_z is kept constant, but more to the point is this interpretation: for a given energy transfer, the axial field component, E_z, decreases in magnitude with the square of frequency. The transverse field components approach a constant value. This sounds reasonable, for consider: as the frequency approaches infinity more and more wavelengths can exist over

the cross section. The phase velocity goes to that of light in free space. It is natural that the wave in the center of the guide notices less and less of the boundary condition; E_z approaches zero and the wave farthest from the conducting boundaries, i.e., along the axis, should look more and more like a plane wave in free space. This general reasoning is clarified by reference again to the specific example of parallel planes (Art. 8.07 and Fig. 8.07) where the approach to a plane wave propagating down the guide as frequency increases was especially easy to predict.

8.15 Attenuation Due to Imperfect Conducting Boundaries

If the conducting boundaries are not perfect, a rigorous analysis requires a solution of the differential equations in the metal guides as well as in the dielectric, with a matching of tangential field components at the boundary between the metal and the dielectric. This technique was used in the simple plane case; there we found that if the conductors are good conductors, it is necessary to revise only slightly the results based upon perfect conductors. Attenuation is the major correction. This may be approximated very closely by assuming that currents in the actual conductors are the same as those in the perfect conductors. This current density is obtainable from the magnetic field at the surface of the conductor. By Eq. 8.14(2),

$$|H_T| = \frac{\omega\epsilon_1}{k_c^2}\,|\nabla E_z|$$

Since E_z is zero at all points along the surface of the conductor, there is no tangential derivative of E_z; ∇E_z consists merely of the derivative normal to the conductor.

$$|H_T|_{\text{conductor}} = \frac{\omega\epsilon_1}{k_c^2}\left|\frac{\partial E_z}{\partial n}\right|_{\text{conductor}} \qquad [1]$$

This gives the magnitude of the current density in the axial direction.

$$|J_z| = |H_T|_{\text{conductor}}$$

If the conductor has a skin effect surface resistivity R_s per square, the above current density will produce losses

$$W_L = \frac{R_s}{2}\oint |J_z|^2 dl \quad \text{watts/meter}$$

The integral is taken around the contour of the conductor.

$$W_L = \frac{R_s\omega^2\epsilon_1^2}{2k_c^4}\oint\left[\frac{\partial E_z}{\partial n}\right]^2 dl$$

$$= \frac{R_s}{2\eta_1^2 k_c^2}\left(\frac{f}{f_c}\right)^2 \oint\left[\frac{\partial E_z}{\partial n}\right]^2 dl \qquad [2]$$

The attenuation is approximately the quotient of power loss per unit length by twice the power transfer. Power transfer is given by Eq. 8.14(3).

$$\alpha = \frac{W_L}{2W_T}$$

$$= \frac{R_s}{2Z_{TM}k_c^2}\left[\frac{\oint\left[\frac{\partial E_z}{\partial n}\right]^2 dl}{\int_{\text{c.s.}} E_z^2 dS}\right] \qquad [3]$$

8.16 Attenuation Due to Imperfect Dielectric

If the dielectric is imperfect, it is only necessary to replace the dielectric constant ϵ_1 wherever it appears in the equations by the quantity $\epsilon_1\left(1 - j\frac{\epsilon''}{\epsilon'}\right)$ where

$$\frac{\epsilon''}{\epsilon'} = \frac{\sigma_1}{\omega\epsilon_1}$$

ϵ'' is the loss factor, ϵ' the dielectric constant on the basis of air as unity. The ratio is approximately the power factor (Art. 7.17). If the dielectric is fairly good, the above ratio will be small,

$$\frac{\epsilon''}{\epsilon'} \ll 1$$

The major correction to previous results is the attenuation. If the above value of dielectric constant is substituted in the equation for propagation constant, Eq. 8.12(2),

$$\gamma = \sqrt{k_c^2 - \omega^2\mu_1\epsilon_1\left(1 - j\frac{\epsilon''}{\epsilon'}\right)}$$

If operation is not so near cut-off that $1 - (f_c/f)^2$ is also very small, the

above may be approximated by the first term in the series expansion when ϵ''/ϵ' is small.

$$\dot{\gamma} \cong jk_1\sqrt{1 - (f_c/f)^2}\left[1 - \frac{j\dfrac{\epsilon''}{\epsilon'}}{2[1 - (f_c/f)^2]}\right]$$

The velocity of propagation is thus unchanged in the first approximation, and the attenuation is the real part of the above:

$$\alpha = \frac{k_1 \dfrac{\epsilon''}{\epsilon'}}{2\sqrt{1 - (f_c/f)^2}} = \frac{\sigma_1\eta_1}{2\sqrt{1 - (f_c/f)^2}} \qquad [1]$$

$$k_1 = \omega\sqrt{\mu_1\epsilon_1} = \frac{2\pi}{\lambda_1}$$

It is especially interesting to see that the attenuation due to losses in the dielectric is of the same form for all shapes of guides. That is, the guide and wave type influence this attenuation only as they enter in determining the cut-off frequency, and thereby the ratio of (f_c/f).

There is, of course, a small propagation below cut-off due to an imperfect dielectric, meaning mainly that there is a small change of phase along the guide below cut-off as compared with the dissipationless case in which there is no phase change below cut-off. The characteristic wave impedance will also have a small real component below cut-off, a small imaginary component above cut-off.

8.17 Transverse Electric or *H* Waves

Let us now study the possible waves that may be propagated with a magnetic field component, but no electric field component, in the direction of propagation. Because of the presence of an axial magnetic field component, they are often known as *H waves*. Since electric field is entirely in the plane transverse to direction of propagation, they are also called *transverse electric* (*TE*) waves. The study of these waves is so similar to that of the transverse magnetic waves that much of the detail included in the past articles will be omitted here. Before starting the analysis, note that a completely transverse electric wave can exist in most of the guides to be studied *only if conductors are perfect.* This is true since there is in most of the guides a longitudinal component of current, and this will consequently require a longitudinal component of electric field if conductivity of the conductors is finite. However, it will represent only a small, and usually completely negligible, correction to the

transverse electric waves to be studied, of the same order as that correction found when a transverse electromagnetic wave was modified by finite conductivity.

Since H_z is to be present, it must satisfy the wave equation. It may be written in the form of Eq. 8.08(12):

$$\nabla_{xy}^2 H_z = -(\gamma^2 + k_1^2)H_z = -k_c^2 H_z \tag{1}$$

The remaining field components may be written in terms of H_z by setting $E_z = 0$ in Eqs. 8.08(7) to 8.08(10).

$$\begin{aligned} H_x &= -\frac{\gamma}{k_c^2}\frac{\partial H_z}{\partial x} & E_x &= -\frac{j\omega\mu_1}{k_c^2}\frac{\partial H_z}{\partial y} \\ H_y &= -\frac{\gamma}{k_c^2}\frac{\partial H_z}{\partial y} & E_y &= \frac{j\omega\mu_1}{k_c^2}\frac{\partial H_z}{\partial x} \end{aligned} \tag{2}$$

From these,

$$E_x = \frac{j\omega\mu_1}{\gamma} H_y \qquad E_y = -\frac{j\omega\mu_1}{\gamma} H_x \tag{3}$$

It again follows that electric and magnetic field components are normal to each other, and that lines of constant H_z correspond to electric field lines (see Prob. 8.11). Since electric field lines must enter the surfaces of perfect conductors normally, so must the lines of constant H_z. The required boundary condition at the surface of the perfect conductors may then be satisfied by requiring that the normal derivative of H_z be zero at that surface.

$$\frac{\partial H_z}{\partial n} = 0 \quad \text{at conducting surfaces}$$

With this requirement, there may be derived an equation similar to Eq. 8.12(1).

$$\int_{\text{c.s.}} [\nabla H_z]^2 dS = k_c^2 \int_{\text{c.s.}} H_z^2 dS \tag{4}$$

For plane waves, no phase differences are brought in through derivatives in the transverse plane, so $[\nabla H_z]^2$ and H_z^2 are necessarily positive, and k_c^2 for these waves is again real and positive. It follows that the propagation constant passes through some cut-off point below which there is attenuation only, and above which there is propagation with a real velocity and no attenuation. The forms for cut-off frequency, attenuation below cut-off and phase and group velocities above cut-off are exactly the same as Eqs. 8.12(3) to 8.12(8).

The characteristic wave impedance is somewhat different from that for transverse magnetic waves; as seen from (3),

$$Z_{TE} = \frac{E_T}{H_T} = \sqrt{\frac{E_x^2 + E_y^2}{H_x^2 + H_y^2}} = \frac{j\omega\mu_1}{\gamma} \tag{5}$$

Note that this is the ratio of E_x/H_y and $-E_y/H_x$ for a positively traveling wave; for a negatively traveling wave it is $-E_x/H_y$ and $+E_y/H_x$. It may also be written

$$Z_{TE} = \frac{\eta_1}{\sqrt{1 - (f_c/f)^2}} \tag{6}$$

This impedance is imaginary below cut-off, is infinite at cut-off frequency, and is purely real above cut-off frequency, again emphasizing the point that energy can be propagated along the ideal guide only at frequencies above cut-off.

The amount of power transfer above cut-off when $\gamma = j\beta$ is again obtained by integrating over the cross section the axial component of the Poynting vector which is given by the transverse field components.

Time average $\qquad P_z = \dfrac{E_T H_T}{2} = \dfrac{Z_{TE} H_T^2}{2}$

But

$$|H_T|^2 = |H_x^2 + H_y^2|$$

$$= \frac{\beta^2}{k_c^4}\left[\left(\frac{\partial H_z}{\partial x}\right)^2 + \left(\frac{\partial H_z}{\partial y}\right)^2\right] = \frac{\beta^2}{k_c^4}[\nabla H_z]^2 \tag{7}$$

Power transfer

$$W_T = \int_{\text{c.s.}} P_z dS = \frac{Z_{TE}\beta^2}{2k_c^4}\int_{\text{c.s.}} [\nabla H_z]^2 dS$$

By the expression for Eq. 8.12(6) and the equivalence of (4),

$$W_T = \frac{\eta_1^2 (f/f_c)^2}{2Z_{TE}}\int_{\text{c.s.}} H_z^2 dS \tag{8}$$

The current flow in the conducting boundary has two components in the TE wave. There is an axial component of current due to the transverse component of magnetic field, just as in the transverse magnetic

wave, but there is in addition a transverse or circulating current due to the axial magnetic field.

$$|J_z| = |H_T| = \frac{\beta}{k_c^2}|\nabla H_z|$$

$$|J_T| = |H_z|_{\text{conductor}}$$

Since the boundary condition requires the vanishing of the normal derivative of H_z at the boundary, ∇H_z becomes simply the tangential derivative, $\partial H_z/\partial l$.

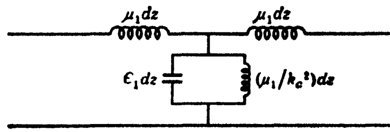


FIG. 8.17. Equivalent circuit for the transverse electric wave.

The power loss per unit length in a real conductor with surface resistivity R_s, carrying the above current, is

$$W_L = \oint \frac{R_s}{2}[|J_z|^2 + |J_T|^2]dl$$

$$= \frac{R_s}{2}\oint\left\{H_z^2 + (f/f_c)^2\frac{[1-(f_c/f)^2]}{k_c^2}\left[\frac{\partial H_z}{\partial l}\right]^2\right\}dl \qquad [9]$$

Attenuation due to this loss is

$$\alpha = \frac{R_s Z_{TE}}{2\eta_1^2}\,\frac{\oint\left\{(f_c/f)^2H_z^2 + 1/k_c^2[1-(f_c/f)^2]\left[\frac{\partial H_z}{\partial l}\right]^2\right\}dl}{\int_{\text{c.s.}} H_z^2 dS} \qquad [10]$$

Attenuation due to imperfect dielectrics is of exactly the same form as for the transverse magnetic waves, since the propagation constant has the same form, and attenuation is obtained by substituting $\epsilon_1\left(1 - j\frac{\epsilon''}{\epsilon'}\right)$ for ϵ_1 in the equation for γ.

Problem 8.17. As in Prob. 8.13, show that the equivalent circuit for transverse electric waves in terms of distributed constants is as pictured in Fig. 8.17.

8.18 Summary of General Results for *TM* and *TE* Waves

Below, in summary form, are the results obtained in the previous articles. They are particularly useful in showing the many simple forms that apply to all shapes of guides.

1. Solve $\nabla_{xy}^2 E_z = -k_c^2 E_z$ (*TM* wave) Solve $\nabla_{xy}^2 H_z = -k_c^2 H_z$ (*TE* wave)

∇_{xy}^2 is the two-dimensional Laplacian in the transverse plane.

2. Find allowed values of k_c from the boundary conditions

$$E_z = 0 \quad \text{along conducting boundary} \qquad (TM \text{ wave})$$

$$\frac{\partial H_z}{\partial n} = 0 \quad \text{along conducting boundary} \qquad (TE \text{ wave})$$

3. Cut-off.

$$f_c = \frac{k_c}{2\pi\sqrt{\mu_1\epsilon_1}}$$

$$\lambda_c = \frac{2\pi}{k_c}$$

4. Attenuation below cut-off.

$$\alpha = k_c\sqrt{1 - (f/f_c)^2}$$

5. Phase velocity and propagation constant above cut-off.

$$v_p = \frac{v_1}{\sqrt{1 - (f_c/f)^2}} \qquad \gamma = j\frac{\omega}{v_1}\sqrt{1 - (f_c/f)^2}$$

6. Group velocity above cut-off.

$$v_g = v_1\sqrt{1 - (f_c/f)^2}$$

7. Characteristic wave impedance (ohms).

$$Z_{TM} = \eta_1\sqrt{1 - (f_c/f)^2} \qquad Z_{TE} = \frac{\eta_1}{\sqrt{1 - (f_c/f)^2}}$$

8. Power transfer above cut-off (watts).

$$(W_T)_{TM} = \frac{Z_{TM}}{2\eta_1^2}\left(\frac{f}{f_c}\right)^2\int_{\text{c.s.}} E_z^2 dS \qquad (W_T)_{TE} = \frac{\eta_1^2}{2Z_{TE}}\left(\frac{f}{f_c}\right)^2\int_{\text{c.s.}} H_z^2 dS$$

9. Power loss in imperfect conductors above cut-off (watts/meter).

$$(W_L)_{TM} = \frac{R_s}{2\eta_1^2 k_c^2}\left(\frac{f}{f_c}\right)^2\oint\left[\frac{\partial E_z}{\partial n}\right]^2 dl$$

$$(W_L)_{TE} = \frac{R_s}{2}\oint\left\{H_z^2 + \frac{\left(\frac{f}{f_c}\right)^2\left[1-\left(\frac{f_c}{f}\right)^2\right]}{k_c^2}\left[\frac{\partial H_z}{\partial l}\right]^2\right\}dl$$

10. Attenuation due to imperfect conductor.

$$\alpha_c = \frac{W_L}{2W_T}$$

11. Attenuation due to imperfect dielectric.

$$\alpha_d = \frac{k_1\frac{\epsilon''}{\epsilon'}}{2\sqrt{1 - (f_c/f)^2}} = \frac{\sigma_1\eta_1}{2\sqrt{1 - (f_c/f)^2}}$$

The ratios α/k_c below cut-off and v_p/v_1 and v_g/v_1 above cut-off are plotted in Fig. 8.18, applying to all TM and TE wave types in all shapes of closed guides.

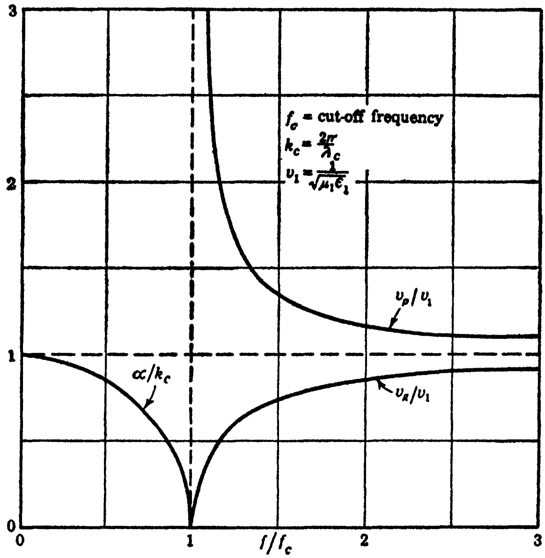


FIG. 8.18. Frequency characteristics of all TE and TM wave types.

8.19 General Wave Types in Rectangular Coordinates

The general solutions for guided waves may be written in rectangular coordinates for application to waves between parallel planes, parallel bar transmission lines, wave guides of rectangular section, etc.

For transverse magnetic or E waves, Eq. 8.18(1) in rectangular coordinates is

$$\nabla_{xy}^2 E_z = \frac{\partial^2 E_z}{\partial x^2} + \frac{\partial^2 E_z}{\partial y^2} = -k_c^2 E_z \qquad [1]$$

This is a partial differential equation which may be solved by the method used in Chapter 3. Assume that the solution may be written as a product of two terms, one a function of x only, the other a function of y only.

$$E_z = XY$$

where X = a function of x only.
Y = a function of y only.

Substitute in (1).

$$X''Y + XY'' = -k_c^2 XY$$

or

$$\frac{X''}{X} + \frac{Y''}{Y} = -k_c^2 \tag{2}$$

The primes indicate derivatives. If this equation is to hold for all values of x and y, since x and y may be changed independently of each other, each of the ratios X''/X and Y''/Y can be only a constant. There are then several forms for the solutions, depending upon whether these ratios are both taken as negative constants, both positive, or one negative and one positive. If both are taken as negative, say k_x^2 and k_y^2 respectively, then,

$$\frac{X''}{X} = -k_x^2$$

$$\frac{Y''}{Y} = -k_y^2$$

The solutions to the above ordinary differential equations are sinusoids, and by (2) the sum of k_x^2 and k_y^2 is k_c^2.

Thus three forms of the wave solution for rectangular coordinates in the transverse plane are listed below, with $e^{j(\omega t - \gamma z)}$ understood. They apply as well to H_z in transverse electric or H waves, since H_z satisfies an equation identical to (1).

$$\left.\begin{matrix} E_z & \text{for } TM \text{ waves} \\ H_z & \text{for } TE \text{ waves} \end{matrix}\right\} = XY \tag{3}$$

where
$$\begin{aligned} X &= A \cos k_x x + B \sin k_x x \\ Y &= C \cos k_y y + D \sin k_y y \end{aligned} \tag{4}$$

$$k_x^2 + k_y^2 = k_c^2$$

or
$$\begin{aligned} X &= A_1 \cos k_x x + B_1 \sin k_x x \\ Y &= C_1 \cosh K_y y + D_1 \sinh K_y y \end{aligned} \tag{5}$$

$$k_x^2 - K_y^2 = k_c^2$$

or
$$\begin{aligned} X &= A_2 \cosh K_x x + B_2 \sinh K_x x \\ Y &= C_2 \cosh K_y y + D_2 \sinh K_y y \end{aligned} \tag{6}$$

$$-(K_x^2 + K_y^2) = k_c^2$$

Note that solutions in the form of (6) have a negative value of k_c^2,

which does not violate previous proofs that k_c^2 must be positive for solutions applying within a closed region, since (6) would not be applicable inside a closed region.

All other components, H_x, H_y, E_x, and E_y are obtained from the above and Eqs. 8.08(7)–(10). For a negatively traveling wave, reverse the sign of all terms containing γ in those equations.

Problem 8.19. Discuss the types of geometrical configurations to which each of the forms of Eqs. 8.19(4)–(6) might be applied.

8.20 General Wave Types in Cylindrical Coordinates

In cylindrical structures, such as coaxial lines or wave guides of circular section, the wave components will be most conveniently expressed in terms of cylindrical coordinates. The two-dimensional Laplacian ∇_{xy}^2 in Eq. 8.18(1), should be written in cylindrical coordinates.

$$\nabla_{xy}^2 E_z = \nabla_{r\phi}^2 E_z = \frac{\partial^2 E_z}{\partial r^2} + \frac{1}{r}\frac{\partial E_z}{\partial r} + \frac{1}{r^2}\frac{\partial^2 E_z}{\partial \phi^2}$$

So that

$$\frac{\partial^2 E_z}{\partial r^2} + \frac{1}{r}\frac{\partial E_z}{\partial r} + \frac{1}{r^2}\frac{\partial^2 E_z}{\partial \phi^2} = -k_c^2 E_z \qquad [1]$$

For this partial differential equation, we shall again substitute an assumed product solution and attempt to separate variables in order to obtain two ordinary differential equations.

Assume

$$E_z = RF_\phi$$

where R = function of r alone.
F_ϕ = function of ϕ alone.

$$R''F_\phi + \frac{R'F_\phi}{r} + \frac{F_\phi''R}{r^2} = -k_c^2 RF_\phi$$

Separating variables,

$$r^2\frac{R''}{R} + \frac{rR'}{R} + k_c^2 r^2 = \frac{-F_\phi''}{F_\phi}$$

The left side of the equation is a function of r alone; the right of ϕ alone. If both sides are to be equal for all values of r and ϕ, both sides must equal a constant. Let this constant be ν^2. There are then the two ordinary differential equations

$$\frac{-F_\phi''}{F_\phi} = \nu^2 \qquad [2]$$

and

$$r^2 \frac{R''}{R} + rR' + k_c^2 r^2 = \nu^2$$

or

$$R'' + \frac{1}{r} R' + \left(k_c^2 - \frac{\nu^2}{r^2}\right) R = 0 \tag{3}$$

The solution to (2) is in sinusoids. By comparing with Eq. 3.18(3), it is seen that solutions to (3) may be written in terms of Bessel functions of order ν. Since H_z for transverse electric or H waves satisfies the same equation as (1), solutions to H_z will also be in the same form. Thus, with $e^{j(\omega t - \gamma z)}$ understood,

$$\left.\begin{array}{ll} E_z & \text{(for } TM \text{ waves)} \\ H_z & \text{(for } TE \text{ waves)} \end{array}\right\} = RF_\phi \tag{4}$$

where

$$\begin{aligned} R &= AJ_\nu(k_c r) + BN_\nu(k_c r) \\ F_\phi &= C \cos \nu\phi + D \sin \nu\phi \end{aligned} \tag{5}$$

or

$$\begin{aligned} R &= A_1 H_\nu^{(1)}(k_c r) + B_1 H_\nu^{(2)}(k_c r) \\ F_\phi &= C \cos \nu\phi + D \sin \nu\phi \end{aligned} \tag{6}$$

or

$$\begin{aligned} R &= A_2 J_\nu(k_c r) + B_2 H_\nu^{(1)}(k_c r) \\ F_\phi &= C \cos \nu\phi + D \sin \nu\phi \end{aligned} \tag{7}$$

The Hankel function form of (6) is useful when it is desired to look at waves as though propagation were in the radial direction, as will be seen in the study of radial transmission lines. The form of (7) is useful for problems in which the constant k_c may be imaginary, since J_ν and $H_\nu^{(1)}$ of imaginary quantities are tabulated.[5]

Other components, E_r, E_ϕ, H_r, and H_ϕ are obtainable from the above solutions by the following equations which are the cylindrical coordinate equivalents of Eqs. 8.08(7)-(10).

$$E_r = -\frac{1}{k_c^2}\left[\gamma \frac{\partial E_z}{\partial r} + \frac{j\omega\mu_1}{r} \frac{\partial H_z}{\partial \phi}\right] \tag{8}$$

$$E_\phi = \frac{1}{k_c^2}\left[-\frac{\gamma}{r} \frac{\partial E_z}{\partial \phi} + j\omega\mu_1 \frac{\partial H_z}{\partial r}\right] \tag{9}$$

[5] See Jahnke-Emde, "Tables of Functions," Dover Publications, 1943.

$$H_r = \frac{1}{k_c^2}\left[\frac{j\omega\epsilon_1}{r}\frac{\partial E_z}{\partial \phi} - \gamma\frac{\partial H_z}{\partial r}\right] \qquad [10]$$

$$H_\phi = -\frac{1}{k_c^2}\left[j\omega\epsilon_1\frac{\partial E_z}{\partial r} + \frac{\gamma}{r}\frac{\partial H_z}{\partial \phi}\right] \qquad [11]$$

For a negatively traveling wave, reverse the sign of all terms containing γ in the above.

Problem 8.20. Demonstrate, making use of the form of Eq. 8.20(7), that a solution with k_c imaginary cannot apply inside a closed region.

8.21 Comparisons of General Wave Behavior and Physical Explanations of Wave Types

Many characteristics have been found in the past articles for waves along uniform guiding systems by mathematical analyses starting from Maxwell's equations. It has been found, for instance, that transverse electromagnetic waves (waves with no field components in the direction of propagation) may propagate along an ideal guide with the velocity of light for the dielectric of the guide. In the transverse plane, these may have any field distributions which correspond to static field distributions. Thus such waves may propagate along a system of two or more conductors, or outside a single conductor, but not inside any hollow pipe, since a static field distribution cannot exist inside an infinitely long hollow, closed conductor. Moreover, it has been verified that the usual transmission line equations written with distributed inductance and capacitance calculated at DC are exact for ideal lines, and the usual equations with distributed inductance, capacitance, resistance, and conductance are excellent approximations for any practical transmission line efficient for energy transfer.

So much for these principal or transmission line waves we have known of, if without assurance, from the conventional line equations, which Maxwell's equations actually verified. In addition, waves have been found which could not have been predicted from the classical transmission line equations based on circuit notions. These waves have either electric or magnetic field components in the direction of propagation. They may propagate inside closed hollow conductors, but only above certain critical or cut-off frequencies for which cross-sectional dimensions between conductors are of the order of a half wavelength. Below these cut-off frequencies the waves, even if started, attenuate extremely rapidly, so that for ordinary transmission lines where spacing between conductors is much smaller than a half wavelength, these waves should not enter into energy propagation. They may be important at discon-

tinuities, end effects, or in the radiation field at a long distance from the line. However, above the cut-off frequency, these waves may be quite satisfactory for energy transfer in any system, and are the only waves which may exist inside closed hollow conductors.

These and other characteristics were obtained by mathematical analysis. It will be profitable to pause now, attempting to understand physically the basis for this behavior and the comparisons between the several types of waves.

It should first be recalled that at the frequencies of interest — at least for the profitable use of hollow pipe wave guides — current flow in the conducting walls will be completely governed by skin effect. For many purposes the conductors may be considered perfect so that there is no penetration whatever into the conductors, but all currents and charges reside on the surface. Even when actual conductivities of practical conducting materials are taken into account, it is found that at such frequencies depth of penetration is of the order of 10^{-4} inch, and the outside of the pipe is perfectly shielded from the fields which are being retained on the interior.

For the dielectric space inside the pipe, it should be recalled:

1. Electric field lines may begin and end on charges. If an electric field ends on a conductor it must represent a charge induced on that conductor.

2. Magnetic field lines can never end since magnetic charges are not known physically. Magnetic fields must always form continuous closed paths, surrounding either a conduction current or a changing electric field (displacement current).

3. Electric field lines may form continuous closed paths, surrounding a changing magnetic field.

In a transverse electromagnetic field, by definition, there are no axial field components; both electric and magnetic fields must lie in the transverse plane. Since electric field is transverse, it would be impossible for magnetic field to surround it without having a component in the axial direction. Consequently all magnetic fields must surround axial conduction currents and not displacement currents. This is the result checked by the analysis for these waves and explains physically why the magnetic fields satisfy a Laplacian equation in the transverse plane outside of the current-carrying region. Similarly, since magnetic fields are transverse, electric fields could not enclose them without having an axial component of electric field. Consequently, in a given transverse plane, all electric field lines must begin on a certain number of positive charges and end on the same number of negative charges. So electric

field also must satisfy Laplace's equation in the transverse plane for the region between conductors.

We can also see quite easily that there can be no transverse electromagnetic waves inside hollow closed conductors. Consider, for a specific case, the round hollow pipe of Fig. 8.21a. If the conductor of the pipe is perfect, magnetic field must be tangential to the conductor. Since magnetic field must also form closed lines, any magnetic field line just

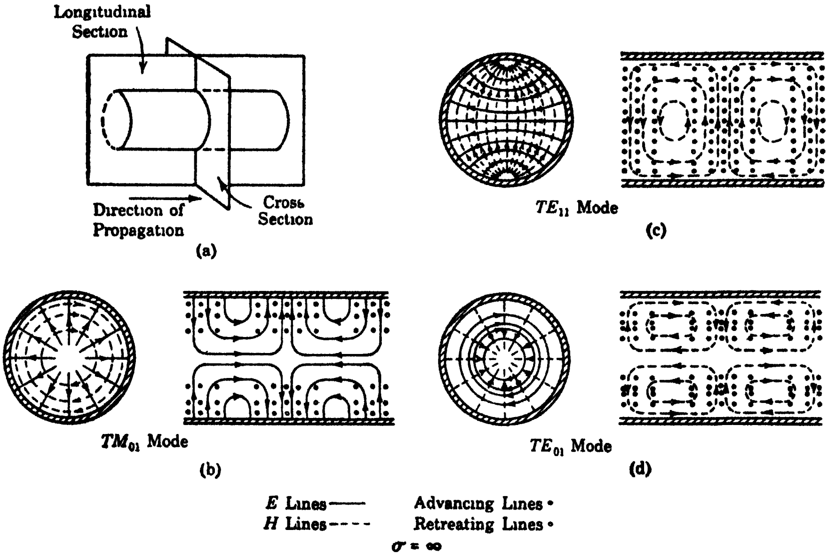


FIG. 8.21. Field distributions for some waves in a hollow circular cylinder.

inside the pipe would have to be a closed circle tangential to the pipe. It could cut no part of the conductor, and so could surround no conduction current. For a wave with no axial electric field, it cannot surround displacement current (or changing electric field). Consequently, it cannot exist at all.

It is evident as an extension of the above reasoning that there may be a value of magnetic field inside the pipe if there is an axial electric field, since the axial displacement current could then account for magnetic field. The electric field might start from positive charges at one section of the guide, turn and go down the guide axially, and finally end on negative charges farther down the guide. (Fig. 8.21b.) It is recognized that such a wave is a transverse magnetic or E wave analyzed in Arts. 8.11 to 8.16. (The subscript notation will be defined in Chapter 9.) Note particularly that since the line integral of magnetic field is propor-

tional to *rate of change* of electric flux enclosed, magnetic field for a single traveling wave is a maximum, not in the plane where axial electric field is a maximum, but rather in the plane where rate of change of axial electric field is a maximum as the entire pattern moves down the guide. If this wave is symmetrical, there must be only axial current flow, produced by the transverse magnetic field at the conductor surface. This is also evident by the current which must flow to account for the lumps of induced charge. From still another point of view, we have agreed that the conducting wall acts as a perfect shield so that no magnetic field can exist outside it due to influences on the inside. Thus, at any section, there must flow a current in the conductor exactly equal and opposite to the axial displacement current inside the guide at that section.

Let us now consider the field distribution for waves with axial magnetic field and transverse electric field. First, as in Fig. 8.21*c*, notice that if electric field lines start from positive charges on one side of the hollow pipe and go directly across to negative charges on the opposite side, magnetic field lines may exist inside the hollow pipe if they surround these electric field lines. In this type of wave there must exist currents flowing circumferentially between the positive and negative charges at any given section in addition to those which flow axially. The former are accounted for by the axial magnetic field at the surface of the conductor; the latter are accounted for by the transverse component of magnetic field at the surface of the conductor.

The wave described above is, of course, a transverse electric or *H* wave. However, another wave of this same type may appear if the electric field lines in the transverse plane do not end on any charges, but always close upon themselves. In this wave (Fig. 8.21*d*) the electric field lines and the magnetic field lines surround each other. There are then no charges induced on the conductors and no axial currents. There are circulating currents arising from the axial component of magnetic field. Since we have found that this axial component becomes very small for frequencies far above cut-off, so will the circulating current become small, and under this condition there will be but slight losses in the guide even though conductors are imperfect. Of course, such a situation indicates that the type of wave described is not so intimately tied to the guide. If it is attempted to make a bend in such a guide, current must flow at the discontinuity, and the new wave generated at the bend may be of an entirely different type. Because of this reason it is often pointed out that the type of wave is unstable. This is the TE_{01} wave of circular guide which will be studied in more detail later.

We might next ask if it is possible to have a transverse magnetic or *E* wave with no charges induced on the guide, but with electric and magnetic fields surrounding each other. A little study of this shows that although it may be possible for the fields to surround each other on the interior of the guide for the higher order *TM* waves, the field nearest the conductor must turn to enter the conductor normally, thus inducing charges as described previously.

All the above general characteristics will be further clarified in later study of the specific waves which may propagate inside guides of circular and rectangular shapes. However, the preceding general study is particularly important in showing that similar types of waves should be found in guides of different cross sections, since the above discussions did not require the specification of the shape of guide.

9

CHARACTERISTICS OF COMMON WAVE GUIDES AND TRANSMISSION LINES

COMMON TRANSMISSION LINES

9.01 Coaxial Lines, Parallel Wire Lines, and Shielded Pairs

From the conclusions of Arts. 8.09 and 8.10 the analysis for ordinary transmission line waves along practical transmission systems may be correctly made from the distributed circuit constant concepts of Chapter 1. For use of the formulas of Chapter 1, it is necessary to calculate values for the inductance, capacitance, resistance, and conductance per unit length. The calculation of such constants was studied in Chapter 6. However, for convenience, some results for the commonly used transmission lines will be listed.

Coaxial lines are among the most commonly used of all transmission lines, particularly at the higher frequencies. This is largely because of the convenient construction and the practically perfect shielding between fields inside and outside of the line. The range of impedances that may be obtained most conveniently by coaxial lines (see Table 9.01) is about 30 to 100 ohms.

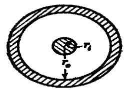
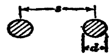
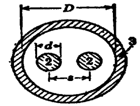
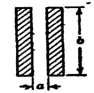
Somewhat higher impedances may be obtained conveniently with parallel wire lines, and these find wide application, although the shielding and radiation problems make them undesirable at the highest frequencies. It is also difficult to attain the lowest impedances conveniently with them. Unlike the coaxial line, the parallel wire line is a balanced line, which is sometimes desirable.

If the parallel wire line is placed inside a conducting pipe as shield, the radiation and shielding difficulties are of course eliminated. The impedance of the line with shield is in general somewhat lower than the same line without the shield. The resulting shielded pair is also a balanced line, assuming symmetrical location of the lines in the shield.

The parallel bar transmission line is sometimes used when balanced lines of low impedance are desired. Like the parallel wire line, it is not perfectly shielded.

In Table 9.01 are listed some of the constants for the above lines. Many of these formulas are approximate, applying at the highest frequencies. For lower frequencies, values of resistance and internal

TABLE 9.01

			$p = \frac{s}{d}$ $q = \frac{s}{D}$ 	 Formulas for $a \ll b$
Capacitance C, farads/meter	$\frac{2\pi\epsilon_1}{\ln\left(\frac{r_0}{r_i}\right)}$	$\frac{\pi\epsilon_1}{\cosh^{-1}\left(\frac{s}{d}\right)}$	....	$\frac{\epsilon_1 b}{a}$
External inductance L, henrys/meter	$\frac{\mu_1}{2\pi}\ln\left(\frac{r_0}{r_i}\right)$	$\frac{\mu_1}{\pi}\cosh^{-1}\left(\frac{s}{d}\right)$	....	$\mu_1\frac{a}{b}$
Conductance G, mhos/meter	$\frac{2\pi\sigma_1}{\ln\left(\frac{r_0}{r_i}\right)} = \frac{2\pi\omega\epsilon_0\epsilon''}{\ln\left(\frac{r_0}{r_i}\right)}$	$\frac{\pi\sigma_1}{\cosh^{-1}\left(\frac{s}{d}\right)} = \frac{\pi\omega\epsilon_0\epsilon''}{\cosh^{-1}\left(\frac{s}{d}\right)}$	....	$\frac{\sigma_1 b}{a} = \frac{\omega\epsilon_0\epsilon' b}{a}$
Resistance R, ohms/meter	$\frac{R_s}{2\pi}\left(\frac{1}{r_0}+\frac{1}{r_i}\right)$	$\frac{2R_s}{\pi d}\left[\frac{s/d}{\sqrt{(s/d)^2-1}}\right]$	$\frac{2R_{s2}}{\pi d}\left[1+\frac{1+2p^2}{4p^4}(1-4q^2)\right] + \frac{8R_{s3}}{\pi D}q^2\left[1+q^2-\frac{1+4p^2}{8p^4}\right]$	$\frac{2R_s}{b}$
Internal inductance L_i, henrys/meter (for high frequency)	← $\frac{R}{\omega}$ →			
Characteristic impedance at high frequency Z_0, ohms	$\frac{\eta_1}{2\pi}\ln\left(\frac{r_0}{r_i}\right)$	$\frac{\eta_1}{\pi}\cosh^{-1}\left(\frac{s}{d}\right)$	$\frac{\eta_1}{\pi}\left\{\ln\left[2p\left(\frac{1-q^2}{1+q^2}\right)\right] - \frac{1+4p^2}{16p^4}(1-4q^2)\right\}$	$\eta_1\frac{a}{b}$
Z_0 for air dielectric	$60\ln\left(\frac{r_0}{r_i}\right)$	$120\cosh^{-1}\left(\frac{s}{d}\right) \cong 120\ln\left(\frac{2s}{d}\right)$ if $s/d \gg 1$	$120\left\{\ln\left[2p\frac{(1-q^2)}{(1+q^2)}\right] - \frac{1+4p^2}{16p^4}(1-4q^2)\right\}$	$120\pi\frac{a}{b}$
Attenuation due to conductor α_c	← $\frac{R}{2Z_0}$ →			
Attenuation due to dielectric α_d	← $\frac{GZ_0}{2} = \frac{\sigma_1\eta_1}{2} = \frac{\pi\sqrt{\epsilon_1'\mu_1'}}{\lambda_0}\left(\frac{\epsilon_1''}{\epsilon_1'}\right)$ →			
Total attenuation db/meter	← $8.686(\alpha_c+\alpha_d)$ →			
Phase constant for low-loss lines β	← $\omega\sqrt{\mu_1\epsilon_1} = \frac{2\pi}{\lambda_1}$ →			

All units above are mks.

$\epsilon_1 = \epsilon_1'\epsilon_0$ = dielectric constant, farads/meter } for the dielectric
$\mu_1 = \mu_1'\mu_0$ = permeability, henrys/meter } for the dielectric
$\eta_1 = \sqrt{\mu_1/\epsilon_1}$ ohms

ϵ_1'' = loss factor of dielectric $= \sigma_1/\omega\epsilon_0$
R_s = skin effect surface resistivity of conductor, ohms
λ_1 = wavelength in dielectric $= \lambda_0/\sqrt{\epsilon_1'\mu_1'}$

Formulas for shielded pair obtained from Green, Leibe, and Curtis, *Bell System Tech. Journ.*, 15, pp. 248–284 (April, 1936).

inductance should be calculated by the methods of Chapter 6 and substituted in the formulas of Chapter 1.

$$\gamma = \alpha + j\beta = \sqrt{(R + j\omega L)(G + j\omega C)}$$

$$Z_0 = \sqrt{\frac{(R + j\omega L)}{(G + j\omega C)}} \text{ ohms}$$

9.02 Coaxial Lines — Higher Order Waves

In addition to the ordinary transmission line wave in a coaxial line, there may exist under certain conditions higher order waves with electric or magnetic field in the direction of the line axis. Such waves would be expected from the study of the simple case of parallel planes, and by the general study of waves along uniform systems in Chapter 8, where *TM* and *TE* waves were found in addition to the principal or transmission line waves. The general forms for the *TM* and *TE* waves in cylindrical coordinates are listed in Art. 8.20. The boundary conditions require that E_z for the *TM* waves be zero at the inner radius and, at the outer radius, assuming perfect conductors. (These, of course, refer to radii measured at the boundary between conductors and dielectric, as in Fig. 9.02*a*.)

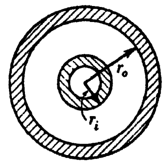


FIG. 9.02*a*. Cross section of a coaxial line.

TM waves

$$A_n J_n(k_c r_i) + B_n N_n(k_c r_i) = 0$$
$$A_n J_n(k_c r_0) + B_n N_n(k_c r_0) = 0$$

or

$$\frac{N_n(k_c r_i)}{J_n(k_c r_i)} = \frac{N_n(k_c r_0)}{J_n(k_c r_0)} \tag{1}$$

For *TE* waves, the derivative of H_z normal to the two conductors must be zero at the inner and outer radii.

$$C_n J_n'(k_c r_i) + D_n N_n'(k_c r_i) = 0$$
$$C_n J_n'(k_c r_0) + D_n N_n'(k_c r_0) = 0$$

or

$$\frac{N_n'(k_c r_i)}{J_n'(k_c r_i)} = \frac{N_n'(k_c r_0)}{J_n'(k_c r_0)} \tag{2}$$

Solutions to the transcendental equations (1) and (2) determine the values of k_c for any wave type and any particular values of r_i and r_0. For any wave there is a cut-off frequency determined by the value of k_c.

$$f_c = \frac{k_c}{2\pi\sqrt{\mu_1 \epsilon_1}} \quad \text{or} \quad \lambda_c = \frac{2\pi}{k_c} \tag{3}$$

Solution of the transcendental equations is not simple; cut-off wavelengths for some of the *TM* waves are given in Fig. 9.02*b*. If the radius of curvature is large, the criterion developed for cut-off of higher order waves between parallel planes (Art. 8.07) might well be used as a first

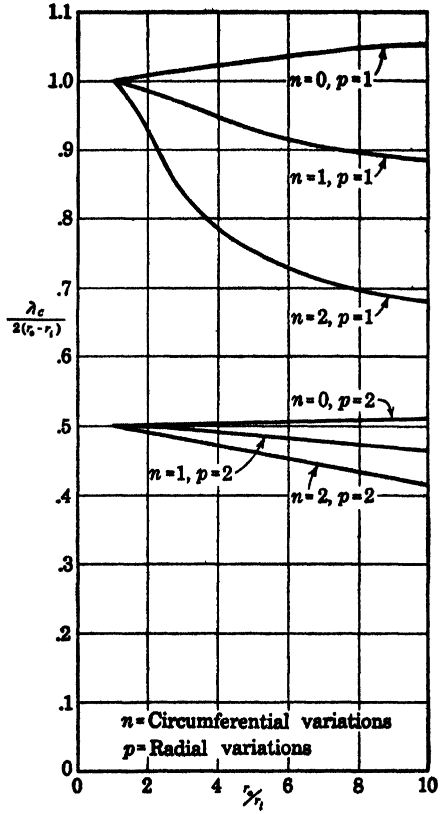


FIG. 9.02*b*. Cut-off wavelengths for some higher order *TM* waves in coaxial lines.

approximation. That is, we expect to find the cut-off of the lowest order *TM* wave when the spacing between conductors is of the order of a half wavelength in the dielectric of the line, and for the *p*th order when it is *p* half wavelengths. That is,

$$\lambda_c \cong \frac{2}{p}(r_0 - r_i) \quad p = 1, 2, 3 \cdots \qquad [4]$$

This is verified by Fig. 9.02*b* for values of r_0/r_i near unity.

Probably more important is the lowest order *TE* wave with circumferential variations. This is analogous to the TE_{10} wave of a rectangu-

lar wave guide, and physical reasoning from the analogy leads one to expect cut-off for this wave type when the average circumference is about equal to wavelength. (The later discussion of Art. 9.05 will indicate this clearly.) Solution of (2) reveals this simple rule to be within about 4 per cent accuracy for r_0/r_i up to 5. In general, for the nth order TE wave with circumferential variations,

$$\lambda_c \cong \frac{2\pi}{n}\left(\frac{r_0 + r_i}{2}\right) \quad n = 1, 2, 3 \cdots \qquad [5]$$

There are, of course, other TE waves with further radial variations, and the lowest order of these has a cut-off about the same as the lowest order TM wave.

Once cut-off is found by solution of (1) and (2) or the above approximations, propagation characteristics are determined by the expressions of Art. 8.18. Of course, for the majority of coaxial line applications, dimensions are small enough compared with wavelength so that the waves are far below cut-off. They then do not propagate energy, but attenuate rapidly so that they are important only at end effects, discontinuities, or in the radiation field. For microwave applications, however, the line size may sometimes be large enough to propagate the circumferential mode determined by $n = 1$ in (5).

COMMON WAVE GUIDES

9.03 Wave Guides of Circular Cross Section

If a hollow round pipe with no inner conductor is now considered as a system for guiding electromagnetic energy, it is known from previous discussions that no transverse electromagnetic wave, the principal wave of ordinary lines, may propagate in such a guide. However, there is a large number of possible waves with either electric or magnetic field in the direction of propagation. The basis for analysis of these has been laid down in Arts. 8.11 to 8.18. A physical discussion of the various wave types has been given in Art. 8.21. It now remains to set down the important quantitative relations for guides of circular cross section.

For a circular guide, cylindrical coordinates will, of course, be selected so that the appropriate solutions for the waves may be taken directly from Art. 8.20. There can be no term in $N_n(k_c r)$ since the solution must in this case apply at the origin, $r = 0$ and $N_n(0) = \infty$. For TM waves, E_z is then given by Eqs. 8.20(4) and 8.20(5) with $B = 0$. For TE waves, H_z is given by a like expression. Other field components for the two types of waves follow from Eqs. 8.20(8) to 8.20(11) respectively. General solutions for the two types of waves are then as follows.

Transverse Magnetic or E Waves

$$E_z = AJ_n(k_c r)\begin{Bmatrix}\cos n\phi \\ \sin n\phi\end{Bmatrix}$$

$$H_r = -j\frac{nf}{k_c\eta_1 r f_c} AJ_n(k_c r)\begin{Bmatrix}\sin n\phi \\ -\cos n\phi\end{Bmatrix}$$

$$H_\phi = -j\frac{f}{f_c\eta_1} AJ'_n(k_c r)\begin{Bmatrix}\cos n\phi \\ \sin n\phi\end{Bmatrix} \qquad [1]$$

$$E_\phi = -H_r Z_{TM}$$

$$E_r = H_\phi Z_{TM}$$

Transverse Electric or H Waves

$$H_z = BJ_n(k_c r)\begin{Bmatrix}\cos n\phi \\ \sin n\phi\end{Bmatrix}$$

$$E_r = j\frac{n\eta_1 f}{k_c r f_c} BJ_n(k_c r)\begin{Bmatrix}\sin n\phi \\ -\cos n\phi\end{Bmatrix}$$

$$E_\phi = j\eta_1\frac{f}{f_c} BJ'_n(k_c r)\begin{Bmatrix}\cos n\phi \\ \sin n\phi\end{Bmatrix} \qquad [2]$$

$$H_\phi = \frac{E_r}{Z_{TE}}$$

$$H_r = -\frac{E_\phi}{Z_{TE}}$$

In all the above expressions $e^{j\omega t-\gamma z}$ is understood and γ, Z_{TM}, and Z_{TE} are (Art. 8.18):

$$\gamma = j\frac{\omega}{v_1}\sqrt{1-\frac{f_c^2}{f^2}}$$

$$Z_{TM} = \eta_1\sqrt{1-\frac{f_c^2}{f^2}}$$

$$Z_{TE} = \eta_1\frac{1}{\sqrt{1-\frac{f_c^2}{f^2}}}$$

For a negatively traveling wave ($e^{j\omega t+\gamma z}$ understood) (1) and (2) are still completely valid in their present form if only the signs of Z_{TM} and Z_{TE} are changed wherever they appear.

For transverse magnetic waves, the boundary condition of zero electric field tangential to the conducting boundary, $E_z = 0$ at $r = a$, must require that

$$J_n(k_c a) = 0 \qquad [3]$$

Since the Bessel function $J_n(x)$ has an infinite number of values of x for which it becomes zero, (3) may be satisfied by any one of these. That is, if p_{nl} is the lth root of $J_n(x) = 0$, (3) is satisfied if

$$(k_c)_{nl} = \frac{p_{nl}}{a} \qquad [4]$$

Equation (4) defines a doubly infinite set of possible values for k_c, one for each combination of the integers n and l. Each of these combina-

tions defines a particular wave type by equations (1), in general differing from all others in field distributions, cut-off frequencies, and propagation properties. A particular E or transverse magnetic wave corresponding to two integers n and l is denoted by E_{nl} or TM_{nl}. The integer n describes the number of variations circumferentially; the integer l describes the number of variations radially. The cut-off wavelength or frequency for a particular wave type follows from (4).

$$(\lambda_c)_{TM_{nl}} = \frac{2\pi}{k_c} = \frac{2\pi a}{p_{nl}} \qquad [5]$$

$$(f_c)_{TM_{nl}} = \frac{k_c}{2\pi\sqrt{\mu_1\epsilon_1}} = \frac{p_{nl}}{2\pi a\sqrt{\mu_1\epsilon_1}} \qquad [6]$$

The lowest value of p_{nl} is the first root of the zero order Bessel function, $p_{01} = 2.405$, so that this TM_{01} wave has the lowest cut-off wavelength of all transverse magnetic waves in a given circular pipe. From (5), this cut-off wavelength is $2.61a$. Note that this wavelength is measured at velocity of light in the dielectric filling the guide, $1/\sqrt{\mu_1\epsilon_1}$.

For transverse electric or H waves the required boundary condition is that normal derivative of H_z be zero at all conducting surfaces. This requires

$$J_n'(k_c a) = 0 \qquad [7]$$

So that if p_{nl}' is the lth root of $J_n'(x) = 0$, (7) is satisfied by

$$(k_c)_{nl} = \frac{p_{nl}'}{a} \qquad [8]$$

Equation (8) again defines a doubly infinite number of possible TE wave types corresponding to all the possible combinations of the integers n and l, n describing the number of circumferential variations, l the number of radial variations. A particular H or transverse electric wave type is labeled H_{nl} or TE_{nl}. Cut-off wavelength and frequency are

$$(\lambda_c)_{TE_{nl}} = \frac{2\pi}{p_{nl}'}a \qquad [9]$$

$$(f_c)_{TE_{nl}} = \frac{p_{nl}'}{2\pi a\sqrt{\mu_1\epsilon_1}} \qquad [10]$$

The lowest value of p_{nl}' is not p_{01}' but rather p_{11}', which is 1.84, so that the TE_{11} wave has the lowest cut-off frequency of all transverse electric waves in a given diameter of pipe. From (9) this corresponds to a

cut-off wavelength of $3.41a$. This is also a lower frequency of cut-off than that found for the lowest order TM wave in a given size pipe. Stated in another way, the TE_{11} wave of a given frequency will propagate in a pipe only 76.6 per cent as big as that required to support a TM_{01} wave of the same frequency. The field distributions in these two wave types and others of the more important wave types for a circular guide are sketched in Table 9.03.

Thus the cut-off frequency for a given type of wave is determined by the radius of guide a, and by the order (n and l) through (5), (6) or (9), (10). The field distributions for any frequency f are then determined by (1) or (2); the phase and group velocities are determined by Eq. 8.18(5) and Eq. 8.18(6), attenuation constant below cut-off by Eq. 8.18(4), characteristic wave impedance by Eq. 8.18(7). Attenuation due to an imperfect dielectric filling the guide is given for all wave types by Eq. 8.18(11). For one of the more interesting items, attenuation due to imperfect conductors above cut-off, it is necessary to evaluate the integrals of Eqs. 8.18(8)–(10). We shall show how this is done for a TM_{nl} wave below.

For a TM_{nl} wave, the expression for power transfer is obtained by substituting the expression for E_z in (1) in Eq. 8.18(8).

$$W_T = \frac{Z_{TM}}{2\eta_1^2}\left(\frac{f}{f_c}\right)^2 \int_{\text{c.s.}} E_z^2\,dS$$

$$= \frac{\left(\frac{f}{f_c}\right)^2\sqrt{1-\left(\frac{f_c}{f}\right)^2}}{2\eta_1}\int_0^{2\pi}\int_0^a A^2 J_n^2(k_c r)\cos^2(n\phi)\,r\,dr\,d\phi$$

The integral of the $\cos^2$ term gives a value of π. The integral of the Bessel function is evaluated by Eq. 3.22(5).

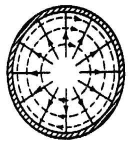
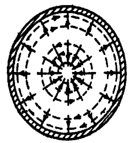
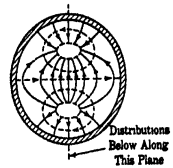
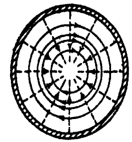
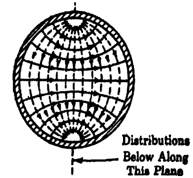
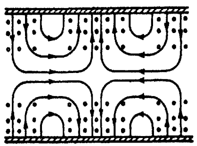
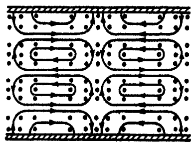
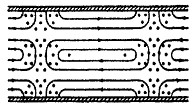
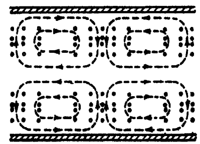
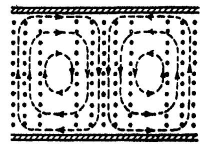
$$\int_0^a J_n^2(k_c r)r\,dr = \frac{a^2}{2}\left[J_n'^2(k_c a) + \left(1 + \frac{n^2}{k_c^2 r^2}\right)J_n^2(k_c a)\right]$$

The second term in this integral is zero because of (3). So

$$W_T = \frac{\pi\left(\frac{f}{f_c}\right)^2\sqrt{1-\left(\frac{f_c}{f}\right)^2}\,a^2}{4\eta_1}A^2 J_n'^2(k_c a) \qquad (11)$$

TABLE 9.03

SUMMARY OF WAVE TYPES FOR CIRCULAR GUIDES

Wave Type	TM_{01}	TM_{02}	TM_{11}	TE_{01}	TE_{11}
Field distributions in cross-sectional plane, at plane of maximum transverse fields					
Field distributions along guide					
Field components present	E_z, E_r, H_ϕ	E_z, E_r, H_ϕ	$E_z, E_r, E_\phi, H_r, H_\phi$	H_z, H_r, E_ϕ	$H_z, H_r, H_\phi, E_r, E_\phi$
p_{nl} or p'_{nl}	2.405	5.52	3.83	3.83	1.84
$(k_c)_{nl}$	$\frac{2.405}{a}$	$\frac{5.52}{a}$	$\frac{3.83}{a}$	$\frac{3.83}{a}$	$\frac{1.84}{a}$
$(\lambda_c)_{nl}$	$2.61a$	$1.14a$	$1.64a$	$1.64a$	$3.41a$
$(f_c)_{nl}$	$\frac{0.383}{a\sqrt{\mu_1\epsilon_1}}$	$\frac{0.877}{a\sqrt{\mu_1\epsilon_1}}$	$\frac{0.609}{a\sqrt{\mu_1\epsilon_1}}$	$\frac{0.609}{a\sqrt{\mu_1\epsilon_1}}$	$\frac{0.293}{a\sqrt{\mu_1\epsilon_1}}$
Attenuation due to imperfect conductors	$\frac{R_s}{a\eta_1}\frac{1}{\sqrt{1-(f_c/f)^2}}$	$\frac{R_s}{a\eta_1}\frac{1}{\sqrt{1-(f_c/f)^2}}$	$\frac{R_s}{a\eta_1}\frac{1}{\sqrt{1-(f_c/f)^2}}$	$\frac{R_s}{a\eta_1}\frac{(f_c/f)^2}{\sqrt{1-(f_c/f)^2}}$	$\frac{R_s}{a\eta_1}\frac{1}{\sqrt{1-(f_c/f)^2}}\left[\left(\frac{f_c}{f}\right)^2+0.420\right]$

The power loss per unit length due to the conductors, by Eq. 8.18(9)

$$W_L = \frac{R_s}{2\eta_1^2 k_c^2}\left(\frac{f}{f_c}\right)^2 \oint \left(\frac{\partial E_z}{\partial n}\right)^2 dl$$

$$= \frac{R_s}{2\eta_1^2 k_c^2}\left(\frac{f}{f_c}\right)^2 k_c^2\, A^2 J_n'^2(k_c a)\int_0^{2\pi} \cos^2(n\phi)\; a d\phi$$

$$= \frac{R_s \pi a}{2\eta_1^2}\left(\frac{f}{f_c}\right)^2 A^2 J_n'^2(k_c a) \qquad [12]$$

By substituting (11) and (12) in Eq. 8.18(10), the attenuation is

$$\alpha_{TM_{nl}} = \frac{W_L}{2W_T} = \frac{R_s}{a\eta_1}\frac{1}{\sqrt{1-\left(\dfrac{f_c}{f}\right)^2}} \quad \text{nepers/meter} \qquad [13]$$

A similar use of the equations gives the attenuation for a TE_{nl} wave,

$$\alpha_{TE_{nl}} = \frac{R_s}{a\eta_1}\frac{1}{\sqrt{1-\left(\dfrac{f_c}{f}\right)^2}}\left[\left(\frac{f_c}{f}\right)^2 + \frac{n^2}{p_{nl}'^2 - n^2}\right] \qquad [14]$$

Some representative curves of attenuation versus diameter are plotted in Fig. 9.03*a* for different wave types at a fixed frequency; and

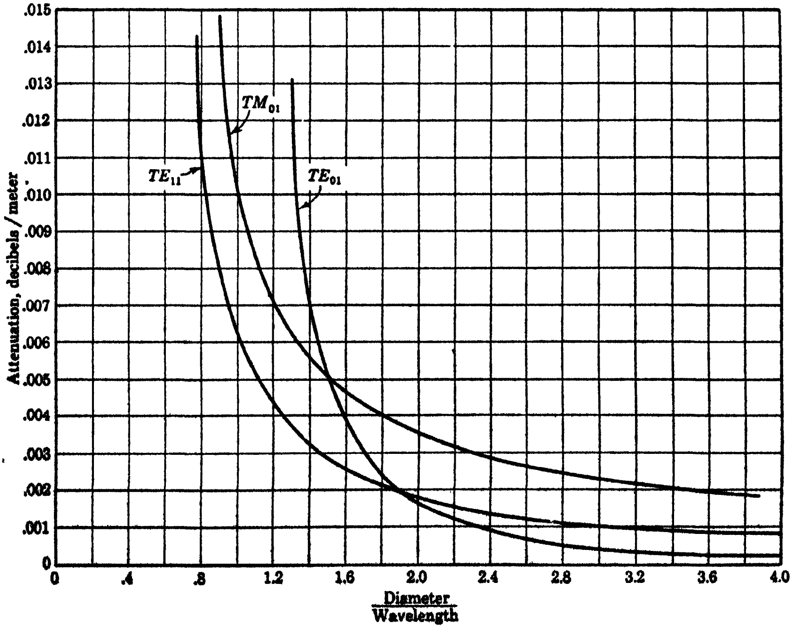


FIG. 9.03*a*. Attenuation due to copper losses in circular wave guides at 3000 mc./sec.

in Fig. 9.03*b*, for different waves types in a guide of fixed diameter, attenuation is plotted versus frequency. The TE_{01} wave is interesting since it shows an attenuation which decreases indefinitely with increasing frequency. This is logical, since equations (2) show that the only magnetic field component tangential to the conductors is H_z, if $n = 0$. As frequency increases, H_z decreases for a constant value of transmitted energy and approaches zero at infinite frequency.

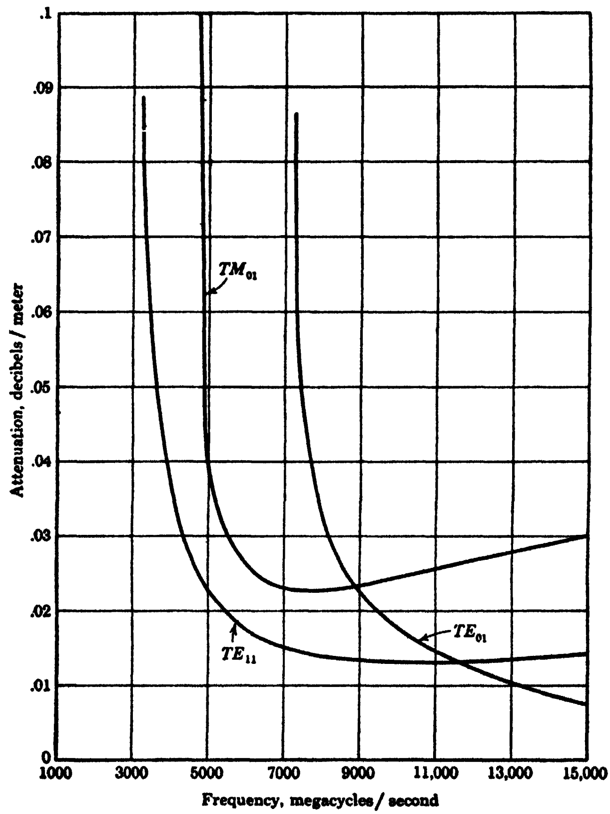


FIG. 9.03*b*. Attenuation due to copper losses in circular wave guides; diameter = 2 in.

Currents in the guide walls therefore approach zero, and losses approach zero. As was pointed out in Art. 8.21, this merely means that under such conditions the wave is not tied intimately to the conducting walls. Any asymmetry or bending of the guide will, of course, produce currents in the walls and a corresponding increase in losses, or may even transform the wave into another type than the TE_{01}. While

the TE_{01} wave was used as an example, all TE_{0l} waves behave similarly.

Problem 9.03(*a*). Derive Eq. 9.03(14) for attenuation of TE_{nl} waves in an imperfectly conducting guide.

Problem 9.03(*b*). Show that for all TM_{nl} waves, the minimum attenuation arising from imperfect conductors occurs at a frequency,

$$f = \sqrt{3} f_c$$

Study the dependence of the value of attenuation at this frequency on f_c. (Recall that R_s is a function of frequency.)

Problem 9.03(*c*). For $\lambda = 7$ cm, select a pipe size to propagate with a reasonable safety factor the TE_{11} wave, but no other wave type. Compare the dissipative attenuation in this TE_{11} wave (copper guide) with the reactive attenuation in the next highest order wave.

9.04 Wave Guides of Rectangular Cross Section

The waves which may propagate inside a hollow pipe guide of rectangular section are similar to those for a guide of circular cross section studied in the last article. The wave types may be written directly from the general results for waves in rectangular coordinates, Art. 8.19.

Transverse Magnetic or E Waves

$$\begin{aligned}
E_z &= A \sin k_x x \sin k_y y \\
H_x &= j\frac{k_y f}{k_c \eta_1 f_c} A \sin k_x x \cos k_y y \\
H_y &= -j\frac{k_x f}{k_c \eta_1 f_c} A \cos k_x x \sin k_y y \\
E_x &= Z_{TM} H_y \\
E_y &= -Z_{TM} H_x
\end{aligned} \qquad [1]$$

Transverse Electric or H Waves

$$\begin{aligned}
H_z &= B \cos k_x x \cos k_y y \\
E_x &= j\frac{\eta_1 k_y f}{k_c f_c} B \cos k_x x \sin k_y y \\
E_y &= -j\frac{\eta_1 k_x f}{k_c f_c} B \sin k_x x \cos k_y y \\
H_x &= -\frac{E_y}{Z_{TE}} \\
H_y &= \frac{E_x}{Z_{TE}}
\end{aligned} \qquad [2]$$

where again $\epsilon^{j\omega t - \gamma z}$ is understood,

$$Z_{TM} = \eta_1 \sqrt{1 - \left(\frac{f_c}{f}\right)^2}$$

$$Z_{TE} = \eta_1 \frac{1}{\sqrt{1 - \left(\frac{f_c}{f}\right)^2}}$$

and

$$\gamma = \sqrt{k_c^2 - k_1^2} = \frac{\omega}{v_1}\sqrt{1 - \left(\frac{f_c}{f}\right)^2}$$

If the wave is a negatively traveling one ($e^{j\omega t+\gamma z}$ understood), (1) and (2) are completely correct except that the sign of Z_{TM} and Z_{TE} should be changed wherever they appear.

The omission of cosine terms in E_z and sine terms in H_z was governed by the boundary requirements that $E_z = 0$ at $x = 0$ and at $y = 0$ for TM waves, and that $\partial H_z/\partial x = 0$ at $x = 0$, $\partial H_z/\partial y = 0$ at $y = 0$ for the TE waves. The requirement of $E_z = 0$ at $x = a$ and $y = b$ also fixes k_x and k_y for TM waves:

$$k_x = \frac{m\pi}{a} \qquad k_y = \frac{n\pi}{b} \tag{3}$$

The requirement of $\partial H_z/\partial x = 0$ at $x = a$ and $\partial H_z/\partial y = 0$ at $y = b$ leads to the same values of k_x and k_y for TE waves. From Eq. 8.19(4) for either TM or TE waves,

$$(k_c)_{m,n} = \sqrt{k_x^2 + k_y^2} = \sqrt{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2} \tag{4}$$

Then cut-off wavelength and frequency may be written

$$(\lambda_c)_{m,n} = \frac{2\pi}{k_c} = \frac{2}{\sqrt{\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2}} = \frac{2ab}{\sqrt{(mb)^2 + (na)^2}} \tag{5}$$

$$(f_c)_{m,n} = \frac{k_c}{2\pi\sqrt{\mu_1\epsilon_1}} = \frac{1}{2\sqrt{\mu_1\epsilon_1}}\sqrt{\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2} \tag{6}$$

There are then a doubly infinite number of possible waves of each type, corresponding to all the combinations of the integers m and n. An E or transverse magnetic wave with m half-sine variations in the x direction and n half-sine variations in the y direction is denoted as an E_{mn} or TM_{mn} wave. An H or transverse electric wave with m half-sine variations in x, n in y, is denoted by H_{mn} or TE_{mn}. Note that by (1) and (2) TE waves may exist with either m or n (but not both) zero, whereas in a TM wave neither m nor n can be zero or the entire wave disappears. The lowest order TE wave, TE_{10}, is of enough special engineering interest to be studied in more detail in a following article. For the moment, however, we see from (5) that the cut-off (free space) wavelength of such a wave is

$$[\lambda_c]_{TE_{10}} = 2a \tag{7}$$

That is, the cut-off frequency is that frequency for which the width of the guide is a half wavelength. It does not depend at all on the

other dimensions. For the lowest order TM, TM_{11},

$$(\lambda_c)_{11} = \frac{2ab}{\sqrt{a^2 + b^2}} \tag{8}$$

For a square, $a = b$, this is merely $\sqrt{2}\,a$.

The phase and group velocities, attenuation below cut-off, and attenuation due to imperfect dielectrics above cut-off for any wave type are given in terms of the cut-off frequency of that wave type by the general expressions of Art. 8.18. For attenuation above cut-off due to imperfect conductivity, we evaluate the integrals of Eqs. 8.18(8)–(10) in a manner similar to that shown for the TM_{mn} wave in a circular guide, Art. 9.03. The results are

$$(\alpha_c)_{TE_{m0}} = \frac{R_s}{b\eta_1\sqrt{1 - \left(\frac{f_c}{f}\right)^2}}\left[1 + \frac{2b}{a}\left(\frac{f_c}{f}\right)^2\right] \tag{9}$$

$$(\alpha_c)_{TE_{mn}} = \frac{2R_s}{b\eta_1\sqrt{1 - \left(\frac{f_c}{f}\right)^2}}\left\{\left(1 + \frac{b}{a}\right)\left(\frac{f_c}{f}\right)^2 + \left[1 - \left(\frac{f_c}{f}\right)^2\right]\left[\frac{\frac{b}{a}\left(\frac{b}{a}m^2 + n^2\right)}{\frac{b^2m^2}{a^2} + n^2}\right]\right\} \tag{10}$$

$$(\alpha_c)_{TM_{mn}} = \frac{2R_s}{b\eta_1\sqrt{1 - \left(\frac{f_c}{f}\right)^2}}\frac{\left[m^2\left(\frac{b}{a}\right)^3 + n^2\right]}{\left[m^2\left(\frac{b}{a}\right)^2 + n^2\right]} \tag{11}$$

Curves of attenuation versus frequency for several representative cases are plotted in Fig. 9.04. Values are again plotted in decibels per meter, this being 8.686 times the value of nepers per meter given in the formulas (9) to (11). There is no wave in a rectangular guide for which attenuation decreases indefinitely with frequency, as there is for the TE_{01} class of waves in the circular guide.

Several field distributions and other data are tabulated in Table 9.04 for the most important waves of a rectangular pipe. It should be especially noted that analogous waves for rectangular and circular pipes are in general not those having the same subscripts. Thus a comparison of these figures with those of Table 9.03 shows that the TE_{10} wave in a rectangular guide is analogous to the TE_{11} wave in a

circular guide; the TM_{11} wave in a rectangular guide is analogous to the TM_{01} wave in a circular guide; the TM_{12} rectangular is analogous to the TM_{11} circular, etc.

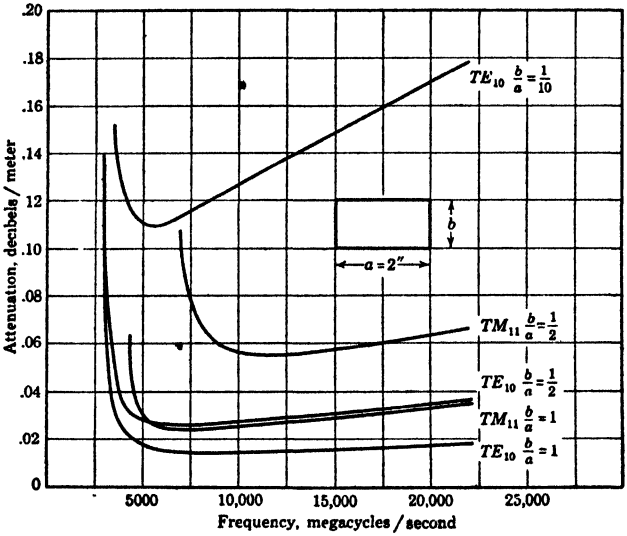


FIG. 9.04. Attenuation due to copper losses in rectangular wave guides of fixed width.

Problem 9.04(*a*). Derive in detail the expressions for attenuation due to imperfect conductors, Eqs. 9.04(9)–(11).

Problem 9.04(*b*). Are there frequencies of minimum attenuation in TM_{mn} and TE_{mn} waves in a rectangular guide as there were for TM_{nl} waves in a circular guide?

Problem 9.04(*c*). Of the wave types studied so far, those transverse magnetic to the axial direction were obtained by setting $H_z = 0$; those transverse electric to the axial direction were obtained by setting $E_z = 0$. For the rectangular wave guide, obtain the lowest order mode with $H_x = 0$ but all other components present. This may be called a wave transverse magnetic to the x direction. Show that it may also be obtained by superposing the TM and TE waves given previously of just sufficient amounts so that H_x from the two waves exactly cancel. Repeat for a wave transverse electric to the x direction. The above wave types are also called *longitudinal section waves.*

9.05 The TE_{10} Wave in a Rectangular Guide

One of the simplest of all the waves which may exist inside hollow pipe wave guides is the TE_{10} wave in the rectangular guide. It is also of great engineering importance, partly for the following reasons.

1. Cut-off frequency is independent of one of the dimensions of the cross section. Consequently for a given frequency this dimension may

be made small enough so that the TE_{10} wave is the only wave which will propagate, and there is no difficulty with higher order waves which end effects or discontinuities may cause to be excited.

2. The polarization of the field is definitely fixed, electric field passing from top to bottom of the guide. This fixed polarization may be required for certain applications.

3. For a given frequency the attenuation due to copper losses is not excessive compared with other wave types in guides of comparable size.

Let us now rewrite the expressions from the previous article for general TE waves in rectangular guides, Fig. 9.05a, setting $m = 1$, $n = 0$ and substituting the value of cut-off for this combination.

$$H_z = B \cos \frac{\pi x}{a} \qquad [1]$$

$$H_x = j\left(\frac{2a}{\lambda_1}\right)\sqrt{1 - \left(\frac{\lambda_1}{2a}\right)^2}\, B \sin \frac{\pi x}{a} \qquad [2]$$

$$E_y = -Z_{TE}H_x \qquad [3]$$

$$H_y = 0 = E_x \qquad [4]$$

$$Z_{TE} = \frac{\eta_1}{\sqrt{1 - \left(\frac{\lambda_1}{2a}\right)^2}} \qquad [5]$$

$$v_p = \frac{1}{\sqrt{\mu_1\epsilon_1}\sqrt{1 - \left(\frac{\lambda_1}{2a}\right)^2}} \qquad [6]$$

$$v_g = \frac{1}{\sqrt{\mu_1\epsilon_1}}\sqrt{1 - \left(\frac{\lambda_1}{2a}\right)^2} \qquad [7]$$

$$\lambda_c = 2a \qquad [8]$$

$$f_c = \frac{1}{2a\sqrt{\mu_1\epsilon_1}} \qquad [9]$$

Attenuation due to imperfect dielectric

$$\alpha_d = \frac{\sigma_1\eta_1}{2\sqrt{1 - \left(\frac{f_c}{f}\right)^2}} = \frac{k_1 \frac{\epsilon_1''}{\epsilon_1'}}{2\sqrt{1 - \left(\frac{f_c}{f}\right)^2}} \qquad [10]$$

Attenuation due to imperfect conductor

$$\alpha_c = \frac{R_s}{b\eta_1\sqrt{1 - \left(\frac{f_c}{f}\right)^2}}\left[1 + \frac{2b}{a}\left(\frac{f_c}{f}\right)^2\right] \qquad [11]$$

In the above, v_p is phase velocity, v_g is group velocity, μ_1, ϵ_1, and η_1 are permeability, dielectric constant, and intrinsic impedance respectively for the dielectric filling the guide. R_s is the skin effect surface resistivity of the conducting walls, and ϵ_1''/ϵ_1' is the power factor of the dielectric.

A study of the field distributions (1) to (3) shows the field patterns for this wave sketched in Table 9.04. First it is noted that no field components vary in the vertical or y direction. The only electric field component is that vertical one E_y passing between top and bottom of the guide. This is a maximum at the center and zero at the conducting walls, varying as a half-sine curve. The corresponding charges induced by the electric field lines ending on conductors are:

(*a*) Charges zero on side walls.

(*b*) A charge distribution on top and bottom corresponding to E_y.

$$\rho_s = -\epsilon_1 E_y \text{ coulombs/meter}^2 \text{ on top}$$
$$= \epsilon_1 E_y \text{ coulombs/meter}^2 \text{ on bottom}$$

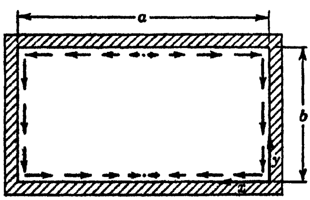


FIG. 9.05*a*. Cross section indicating transverse current flow around guide in TE_{10} wave.

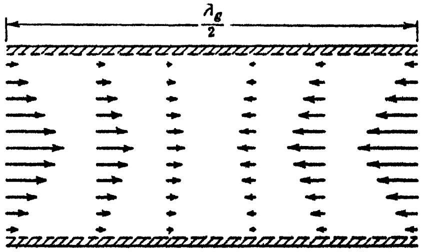


FIG. 9.05*b*. Top view indicating longitudinal current flow along guide in TE_{10} wave.

The magnetic field forms closed paths surrounding the vertical electric displacement currents arising from E_y, so that there are components H_x and H_z. H_x is zero at the two side walls and a maximum in the center, following the distribution of E_y. H_z is a maximum at the side walls and zero at the center. H_x corresponds to a longitudinal current flow down the guide in the top, and opposite in the bottom; H_z corresponds to a current from top to bottom around the periphery of the

guide. These current distributions are sketched in Figs. 9.05*a* and *b*.

(*a*) Longitudinal current flow:

On top $J_z = H_x$ amperes per meter.

On bottom $J_z = -H_x$ amperes per meter.

(*b*) Transverse current flow from top to bottom:

On walls $J_y = -H_z|_{x=0}$ amperes per meter.

On top $J_x = -H_z$ amperes per meter.

On bottom $J_x = H_z$ amperes per meter.

This simple wave type is a convenient one to study in order to strengthen some of our physical pictures of wave propagation. First note that this is one of the types predicted by physical reasoning in Art. 8.21. Electric field is confined to the transverse plane and so passes between equal and opposite charge densities lying on different parts of the walls in the same transverse plane. Currents flow around the periphery of the guide between these opposite charges; currents also flow longitudinally down the guide between a given charge and that of opposite sign, a half wave farther down the guide. The magnetic fields surround the electric displacement currents inside the guide and so must have an axial as well as a transverse component.

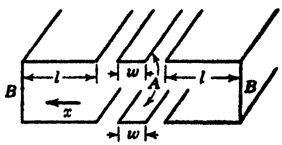


FIG. 9.05*c*.

We might now look at the problem from a little different angle, imagining how an engineer familiar only with transmission line techniques might conceivably arrive at the distribution of fields and currents in such a wave guide. Suppose he first imagines a parallel strip transmission line A, as in Fig. 9.05*c*, with voltage between the two strips, and a going and return current in the two strips. The widths w are small compared with wavelength. Next suppose he wishes to close in the two sides for shielding purposes, recalls that the input impedance of a quarter wave shorted line is infinite, and so decides that he may put closing sections B along the two sides, so long as these are a quarter wavelength along the dimension l. These should then look like a quarter wave shorted line to any currents trying to flow in the direction x, so that there are only infinite impedances connected across the two sides of the parallel strip line, and the ordinary operation of this line should not be interfered with. He has arrived at a minimum overall width for his closed section of $2l + w$, or somewhat greater than a half wavelength, since $l = \lambda/4$. He recognizes, of course, that fields

must penetrate into the two closing side sections and a transverse current flow will exist in these which is a maximum at the shorted ends where voltage is zero, so that all these conclusions correspond to those we have obtained previously. Of course, it is apparent that the fields in the shorting sides and in the original transmission line are not really separated, and the phase velocities along the middle section are changed by the side members, so that this is not a particularly good way to analyze this problem, and it will not be pursued further. However, it should stress the fact that the idea of wave propagation inside a closed wave guide is not at all contrary to ordinary engineering ideas which embrace transmission line techniques, once it is recognized that transverse dimensions of these guides must be comparable with wavelength.

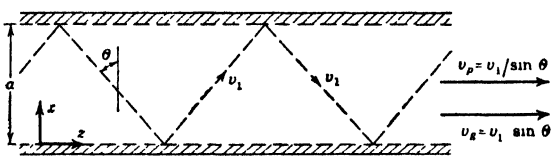


FIG. 9.05*d*. Path of uniform plane wave component of TE_{10} wave in rectangular guide.

A third viewpoint follows from that used in studying the higher order waves between parallel planes. Here it was pointed out that one could visualize the *TM* and *TE* waves in terms of plane waves bouncing between the two planes at such an angle that the interference pattern maintains a zero of electric field tangential to the two planes. Similarly, the TE_{10} wave in the rectangular guide may be thought of as arising from the interference between incident and reflected plane waves, polarized so that the electric vector is vertical, and bouncing between the two sides of the guide at such an angle with the sides that the zero electric field is maintained at the two sides. One such component uniform plane wave is indicated in Fig. 9.05*d*. As in the result of Art. 8.07, when the width a is exactly $\lambda_1/2$, the waves travel exactly back and forth across the guide with no component of propagation in the axial direction. At slightly higher frequencies there is a small angle θ such that $a = \lambda_1/2 \cos\theta$, and there is a small propagation in the axial direction, a very small group velocity in the axial direction $v_1 \sin\theta$, and a very large phase velocity $v_1/\sin\theta$. At frequencies approaching infinity, θ approaches 90°, so that the wave travels down the guide practically as a plane wave in space propagating in the axial direction.

All the above points of view explain why the dimension b should not

enter into the determination of cut-off frequency. Since the electric field is always normal to top and bottom, the placing of these planes plays no part in the boundary condition. However, this dimension b will be important from two other points of view.

(a) The smaller b is (all other parameters constant), the greater is the electric field across the guide for a given power transfer, and so the danger of voltage breakdown is greater.

(b) The smaller b is (all other parameters constant), the greater is the attenuation due to conductor losses.

The first point is easily seen since it was shown that the power transfer can be written as the integral over the cross-sectional area of E^2/Z_{TE}. Z_{TE} does not change with b, so as cross-sectional area decreases, E must increase, if power is to be constant.

The second point follows from an approximate picture in which the attenuation is roughly proportional to the ratio of perimeter to cross-sectional area. This picture is a logical one as the conductor losses occur on the perimeter, and the power transfer occurs through the cross-sectional area. Of course, field distributions enter, and we can look at this case more rigorously by noting that if the strength of magnetic field is maintained constant as b is decreased, the magnitude of currents in the walls is maintained constant. A large part of the losses occur along the top and bottom, and this part is consequently unchanged as b decreases, but power transfer for this constant H decreases directly with b. Therefore the ratio of power loss to power transfer increases as b decreases.

Problem 9.05(*a*). For $\lambda_1 = 10$ cm, design a rectangular wave guide with copper conductor and air dielectric so that the TE_{10} wave will propagate with a 30 per cent safety factor ($f = 1.30 f_c$) but also so that the wave type with next higher cut-off will be 30 per cent below its cut-off frequency.

Calculate the attenuation due to copper losses in decibels per meter.

Problem 9.05(*b*). Repeat the above for $\lambda_1 = 5$ cm.

Problem 9.05(*c*). Design for the same frequency and conditions of (*a*) except that the guide is to be filled with a dielectric having a dielectric constant 4 times that of air. Calculate the increase in attenuation due to copper losses alone, assuming the dielectric is perfect. Calculate the additional attenuation due to this dielectric, if $\epsilon_1''/\epsilon_1' = 0.01$.

Problem 9.05(*d*). Sketch lines showing direction of *total* current flow in the guide walls for a single traveling TE_{10} wave.

OTHER WAVE GUIDING SYSTEMS

9.06 Dielectric Rod or Slab Guides

The study of waves in the rectangular guide from the point of view of plane waves reflected between top and bottom (Art. 9.05) suggests that

under certain conditions a wave may be guided without loss of energy by a slab of perfect dielectric having no metal boundaries. This follows from the concept of total reflection of Art. 7.13, where it was found that if a wave traveling in a dense dielectric strikes the boundary of a less dense dielectric at an angle of incidence greater than a certain critical angle, all energy is reflected. This critical angle, Eq. 7.13(2), where 1 refers to the dense medium and 2 to the less dense medium, is

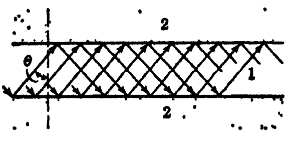


FIG. 9.06a. Paths of uniform plane wave components in a dielectric slab guide.

$$\theta_c = \sin^{-1}\left(\sqrt{\frac{\mu_2\epsilon_2}{\mu_1\epsilon_1}}\right) \qquad [1]$$

Thus in a dielectric slab as in Fig. 9.06a, which is assumed infinite in the direction normal to the paper, suppose that plane waves are excited inside the dielectric in some manner so that they travel as shown, striking the surface at an angle of incidence, θ. If $\theta > \theta_c$ all energy will be reflected at each reflection and all will be retained in the slab. As with other wave guides, there is, for a slab of given thickness, a certain minimum frequency at which such a condition can exist. For frequencies lower than this critical frequency, the angle θ will be less than θ_c and a certain amount of energy will be transmitted into the dielectric medium 2 at each reflection, so that the dielectric does not act as a perfect guide. At frequencies higher than the critical, the angle becomes greater than θ_c, and the only fields in medium 2 are reactive fields that decay exponentially from the boundary in the transverse direction. No average energy is then transmitted into this region. As the frequency approaches infinity, $\theta \to \pi/2$ and the exponentially decaying fields in medium 2 approach zero. The critical frequency is that for which $\theta = \theta_c$. A study of the incident and reflected waves at this critical angle shows that there is a phase angle of 180° between incident and reflected components of magnetic fields parallel to the surface. It follows (maybe not obviously) that the slab should be exactly a half wave thick, measured at a phase velocity transverse to the slab.

$$d = \frac{\lambda_1}{2\cos\theta} = \frac{1}{2f\sqrt{\mu_1\epsilon_1}\cos\theta} \qquad [2]$$

Substitute the value of $\theta = \theta_c$ from (1)

$$\cos\theta_c = \sqrt{1 - \sin^2\theta_c} = \sqrt{1 - (\epsilon_2\mu_2/\mu_1\epsilon_1)}$$

So

$$f_c = \frac{1}{2d\sqrt{\mu_1\epsilon_1}\sqrt{1 - (\epsilon_2\mu_2/\mu_1\epsilon_1)}} = \frac{1}{2d\sqrt{\mu_1\epsilon_1 - \mu_2\epsilon_2}} \qquad [3]$$

Note that if $\mu_1\epsilon_1 \gg \mu_2\epsilon_2$, the requirement for cut-off is that the slab be a half wave thick, measured for the dielectric material of the slab, so that such a slab will have exactly the same cut-off frequency as though it had conducting walls. If $\mu_2\epsilon_2$ is not negligible compared with $\mu_1\epsilon_1$, the guide must be somewhat thicker than a similar slab with conducting boundaries in order to have the same cut-off frequency.

For exact behavior of the guided wave below and above this critical frequency it would be possible to utilize further the results of reflections at boundaries between dielectrics, but probably it would be as easy to go directly to Maxwell's equations and match solutions on the two sides of the boundary. For variety, let us do this, not for the above example but for a round dielectric rod in a medium of lesser dielectric constant.

Let us investigate the possibilities of propagating a *TM* wave with circular symmetry in a dielectric rod (medium 1 of Fig. 9.06*b*) surrounded by the dielectric medium 2 with no intervening conductors. The proper wave solutions may be found from Art. 8.20. If ϕ variations are eliminated, $\nu = 0$. Since medium 1 includes the origin, only the J_0 term can be present in this region; otherwise fields would become infinite at $r = 0$. Since medium 2 extends to infinity, only the $H_0^{(2)}$ term can be present in this solution; otherwise fields would become infinite at $r = \infty$. The factor $e^{(j\omega t - \gamma z)}$ is, of course, understood in all terms. Then,

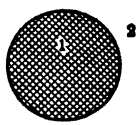


FIG. 9.06*b*. Dielectric rod wave guide.

$$\begin{aligned} E_{z1} &= A_1 J_0(k_{c1} r) \\ E_{z2} &= A_2 H_0^{(2)}(k_{c2} r) \end{aligned} \tag{4}$$

Other components follow from the relations of Eqs. 8.20(8) to 8.20(11):

$$\begin{aligned} E_{\phi 1} &= E_{\phi 2} = 0 & H_{r1} &= H_{r2} = 0 \\ E_{r1} &= \frac{\gamma A_1}{k_{c1}} J_1(k_{c1} r) & E_{r2} &= \frac{\gamma A_2}{k_{c2}} H_1^{(2)}(k_{c2} r) \\ H_{\phi 1} &= \frac{j\omega\epsilon_1 A_1}{k_{c1}} J_1(k_{c1} r) & H_{\phi 2} &= \frac{j\omega\epsilon_2 A_2}{k_{c2}} H_1^{(2)}(k_{c2} r) \end{aligned} \tag{5}$$

where

$$k_{c1}^2 = \gamma^2 + \omega^2 \mu_1 \epsilon_1 \tag{6}$$

$$k_{c2}^2 = \gamma^2 + \omega^2 \mu_2 \epsilon_2 \tag{7}$$

At the boundary between the two dielectrics, $r = a$, E_z and H_ϕ must be

continuous. If this requirement is placed in (4) and (5)

$$\frac{J_0(k_{c1}a)}{J_1(k_{c1}a)} = \frac{\epsilon_1 k_{c2}}{\epsilon_2 k_{c1}} \frac{H_0^{(2)}(k_{c2}a)}{H_1^{(2)}(k_{c2}a)} \tag{8}$$

We now reason as follows. If the condition under which all energy is retained in the rod is sought, and no average energy is to be transmitted into the second medium, it is desired to have a solution corresponding to an exponential decay in the outer medium, and from Chapter 3 we find that this is obtained if k_{c2} is imaginary, since $H_0^{(2)}$ of an imaginary quantity is analogous to a negative exponential. The requirement is then

$$k_{c2}^2 < 0$$

and the critical limiting condition is

$$k_{c2} = 0 \tag{9}$$

From (7), the propagation constant under this critical condition is

$$\gamma^2 = -\omega^2 \mu_2 \epsilon_2$$

$$\gamma = j\beta = j\omega\sqrt{\mu_2\epsilon_2} \tag{10}$$

Therefore, there is propagation with no attenuation and, under this critical condition, at a phase velocity equal to the velocity of light in the outer medium.

If (8) is observed for $k_{c2} = 0$, it is seen that for this critical condition

$$J_0(k_{c1}a) = 0$$

Denote the lth root of $J_0(x) = 0$ by p_{0l}. Then

$$k_{c1}a = p_{0l}$$

But from (6) and (10)

$$k_{c1} = \omega\sqrt{\mu_1\epsilon_1 - \mu_2\epsilon_2}$$

So

$$f_c = \frac{p_{0l}}{2\pi a\sqrt{\mu_1\epsilon_1 - \mu_2\epsilon_2}} \tag{11}$$

The lowest root, p_{01}, is 2.405.

This above value of critical frequency is seen to be quite similar to that of (3), obtained from the wave reflection concept for the slab of dielectric, and for $\mu_1\epsilon_1 \gg \mu_2\epsilon_2$ it reduces to the expression for cut-off frequency of a TM_0 wave in a guide with metal boundaries.

If the above analysis were followed through in detail, it would be discovered that large negative values of k_{c2}^2 correspond to very high

frequencies, and for these the phase velocity approaches $1/\sqrt{\mu_1\epsilon_1}$, or the velocity of light for the medium of the rod. The large imaginary values of k_{c2} require that fields attenuate rapidly as one progresses transversely into the outer dielectric, that is, most of the energy is confined in the rod. Conversely, for the small values of k_{c2} near the critical frequency, the fields extend a long distance into medium 2. These changes in energy distribution check with results from physical reasoning which would lead us to believe that velocity of propagation would be determined largely by region 2 near cut-off, and by region 1 as frequency approaches infinity.

This is, of course, only a partial treatment of the problem for only two of the many waves that may be guided by dielectric rods. However, the preceding physical pictures of total reflection and the analysis based on Maxwell's equations should now supply a definite feeling for some of the fundamental properties of such guided waves.

Problem 9.06(a). Obtain the attenuation due to a lossy dielectric for the TM_{01} wave guided by the circular dielectric rod.

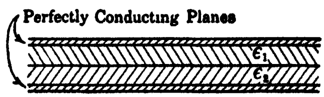


FIG. 9.06c.

Problem 9.06(b). Consider the propagation of waves between two infinite parallel conducting planes separated by two regions of different dielectric constant (Fig. 9.06c). Show that even for perfectly conducting planes, an E_z component must be present in the principal wave and indicate the extent to which the conventional transmission line equations might be in error in predicting the characteristics of the principal wave. Assume spacings small compared with wavelength.

9.07 Waves Guided by a Single Cylindrical Wire

A principal or transverse electromagnetic wave may exist outside a single perfect conductor, since there is a solution to Laplace's equation corresponding to charges on the conductor with electric field lines going to infinity, and a single unidirectional current in the conductor. The proper cross-sectional field distributions outside the wire, if the wire cross section is circular, are

$$E_r = \frac{A}{r} \tag{1}$$

$$H_\phi = \frac{A}{\eta r} \tag{2}$$

All the properties of the general transverse electromagnetic wave found in Art. 8.09 apply, such as the propagation with velocity of light in the surrounding dielectric.

Physically we might think of the wire as the inner conductor of a coaxial transmission line with return at infinity. It is evident that in a practical case the electric field lines will try to end on any other conductors in the vicinity, making these the return if possible. The single wire is consequently not very promising as an ideal transmission system; certainly it is the exact opposite of a well-shielded system. There may be certain cases involving the use of wire conductors in radio-frequency circuits for which results of an analysis for this case are of importance, but the problem is not of enough engineering importance compared to other guiding systems to warrant a complete analysis here. A very excellent and complete analysis of the problem is given by Stratton.[1] Such an analysis shows these interesting points:

1. *TM* or *TE* waves may exist outside the single conductor if the conductor is perfect.

2. If the conductor has finite conductivity, wave solutions inside the conductor may be matched to those outside, and it is found that any higher order waves having variations with circumference have extremely high attenuations.

3. The wave which remains with no circumferential variations reduces to the ideal wave of (1) and (2) as conductivity approaches infinity. For very poor conductors or very small conductor radii, the attenuation in this wave may be terrific, and its phase velocity may depart markedly from the velocity of light in the surrounding dielectric.

4. As the conductivity of the wire becomes poorer, the fields penetrate farther into the conductor. When the conductivity has become very small the wire takes on more the characteristics of a lossy dielectric and the solution approaches the solution for waves guided by a dielectric rod (Art. 9.06).

5. The analysis of this case reminds one that there is a transverse electromagnetic or principal wave possible for the parallel wire line in addition to that already studied. This wave corresponds to like charges on the two wires and equal currents in the same direction in the two lines, with fields extending outward toward infinity. This is the *zero-phase-sequence wave* of power transmission line experience.

9.08 Radial Transmission Lines

Another guide of practical importance consists of two circular, parallel, conducting plates, separated by a dielectric and used for guiding electromagnetic energy radially (Figs. 9.08*a* and *b*). The simplest wave that may be guided by these plates is one with no field

[1] Stratton, "Electromagnetic Theory," McGraw-Hill, 1941.

variations circumferentially or axially. There are then no field components in the radial direction, but field components E_z and H_ϕ only. The component E_z, having no variations in the z direction, corresponds to a total voltage $E_z d$ between plates. The component H_ϕ corresponds to a total radial current $2\pi r H_\phi$, outward in one plate and inward in the other. This wave is then exactly analogous to an ordinary transmission line wave and thus derives its name of radial transmission line.

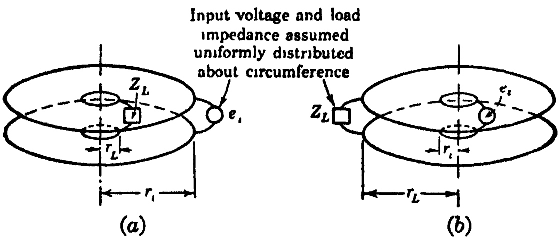


FIG. 9.08. (a) Radial transmission line with input at outer radius. (b) Radial transmission line with input at inner radius.

For the simple wave described above, since there are no radial field components, it is possible to base the analysis on the transmission line equations, except that L and C now vary with radius. However, we already have the wave solutions for fields if results of Art. 8.20 are properly interpreted. Since there are no ϕ variations, ν is set equal to zero. Since there are no z variations, γ is also set equal to zero. In order to identify terms as waves traveling radially inward or radially outward, the form of Eq. 8.20(6) is used. We shall see the reasons for this below. The constant k_c by Eq. 8.11(5) reduces to $k_1 = \omega\sqrt{\mu_1\epsilon_1}$, since $\gamma = 0$.

$$E_z = AH_0^{(1)}(k_1 r) + BH_0^{(2)}(k_1 r) \qquad [1]$$

With γ and $\nu = 0$, the only other remaining field component in Eqs. 8.20(8) to 8.20(11) is H_ϕ.

$$H_\phi = \frac{1}{j\omega\mu_1}\frac{\partial E_z}{\partial r}$$

$$H_\phi = \frac{j}{\eta_1}[AH_1^{(1)}(k_1 r) + BH_1^{(2)}(k_1 r)] \qquad [2]$$

The two terms may be identified definitely as waves traveling inward and outward by employing the asymptotic expressions of the Bessel

functions for large arguments (Art. 3.19). Then

$$E_z\Big|_{k_1r\to\infty} = \sqrt{\frac{2}{\pi k_1 r}}\left[Ae^{j\left(k_1r-\frac{\pi}{4}\right)} + Be^{-j\left(k_1r-\frac{\pi}{4}\right)}\right]$$

with a similar expression for H_ϕ. When the above are multiplied by $e^{j\omega t}$, the first term will involve $e^{j(\omega t+k_1r)}$ and the second $e^{j(\omega t-k_1r)}$ so that these are identified respectively as waves propagating in the negative r and positive r directions.

The wave impedance of an outward traveling wave may be found by taking the ratio of E_z to H_ϕ in (1) and (2) with $A = 0$.

$$Z_r^+ = -\frac{\eta_1}{j}\frac{H_0^{(2)}(k_1r)}{H_1^{(2)}(k_1r)} \qquad [3]$$

This is a function of r. For the inward traveling wave, $B = 0$,

$$Z_r^- = \frac{\eta_1}{j}\frac{H_0^{(1)}(k_1r)}{H_1^{(1)}(k_1r)} \qquad [4]$$

The signs of (3) and (4) are chosen in accordance with the convention discussed in Art. 7.07.

With these definitions of impedance it is possible to evaluate the constants A and B and so find fields at any point along the line if any two field quantities are given, such as a terminating impedance and an electric field, two values of magnetic field, two values of electric field, or one value of electric field and one of magnetic field. Before giving these formulas let us define magnitudes and phase angles for the complex Hankel functions as follows.

$$\begin{aligned}
H_0^{(1)}(x) &= J_0(x) + jN_0(x) = G_0(x)e^{j\theta(x)} \\
H_0^{(2)}(x) &= J_0(x) - jN_0(x) = G_0(x)e^{-j\theta(x)} \\
jH_1^{(1)}(x) &= -N_1(x) + jJ_1(x) = G_1(x)e^{j\psi(x)} \\
jH_1^{(2)}(x) &= -[-N_1(x) - jJ_1(x)] = -G_1(x)e^{-j\psi(x)}
\end{aligned}$$

so that

$$G_0(x) = \sqrt{J_0^2(x) + N_0^2(x)}; \quad \theta(x) = \tan^{-1}\left[\frac{N_0(x)}{J_0(x)}\right]$$

$$G_1(x) = \sqrt{J_1^2(x) + N_1^2(x)}; \quad \psi(x) = \tan^{-1}\left[\frac{J_1(x)}{-N_1(x)}\right]$$

The expressions (1) and (2) then become

$$E_z = G_0(k_1r)[Ae^{j\theta(k_1r)} + Be^{-j\theta(k_1r)}] \qquad [5]$$

$$H_\phi = \frac{G_1(k_1r)}{\eta}[Ae^{j\psi(k_1r)} - Be^{-j\psi(k_1r)}] \qquad [6]$$

The expressions (3) and (4) become

$$Z_r^+ = Z_0(k_1r)e^{j[\psi(k_1r)-\theta(k_1r)]} \qquad [7]$$

$$Z_r^- = Z_0(k_1r)e^{-j[\psi(k_1r)-\theta(k_1r)]} \qquad [8]$$

where

$$Z_0(k_1r) = \eta_1 \frac{G_0(k_1r)}{G_1(k_1r)} \qquad [9]$$

The magnitudes G_0 and G_1, the phase angles θ and ψ, and the impedance Z_0 are plotted in Fig. 9.08c.

The constants A and B will now be determined for several different cases. The resulting formulas are quite similar to the familiar formulas of transmission line theory giving voltages, currents, and impedances in terms of input end or loading end values. In the following, the subscript of a quantity indicates the quantity is to be evaluated at the value of r denoted by the subscript.

1. Given electric field E_a at r_a, magnetic field H_b at r_b; for any radius r,

$$\begin{aligned} E &= E_a \frac{G_0}{G_{0a}} \frac{\cos(\theta - \psi_b)}{\cos(\theta_a - \psi_b)} + jZ_{0b}H_b \frac{G_0}{G_{0b}} \frac{\sin(\theta - \theta_a)}{\cos(\theta_a - \psi_b)} \\ H &= H_b \frac{G_1}{G_{1b}} \frac{\cos(\psi - \theta_a)}{\cos(\theta_a - \psi_b)} + j\frac{E_a}{Z_{0a}} \frac{G_1}{G_{1a}} \frac{\sin(\psi - \psi_b)}{\cos(\theta_a - \psi_b)} \end{aligned} \qquad [10]$$

2. Given electric fields E_a at r_a, E_b at r_b; for any radius r,

$$\begin{aligned} E &= E_a \frac{G_0}{G_{0a}} \frac{\sin(\theta_b - \theta)}{\sin(\theta_b - \theta_a)} + E_b \frac{G_0}{G_{0b}} \frac{\sin(\theta - \theta_a)}{\sin(\theta_b - \theta_a)} \\ H &= \frac{E_b}{jZ_{0b}} \frac{G_1}{G_{1b}} \frac{\cos(\psi - \theta_a)}{\sin(\theta_b - \theta_a)} - \frac{E_aG_1}{jZ_{0a}G_{1a}} \frac{\cos(\theta_b - \psi)}{\sin(\theta_b - \theta_a)} \end{aligned} \qquad [11]$$

3. Given magnetic fields H_a at r_a, H_b at r_b,

$$\begin{aligned} E &= \frac{Z_{0a}H_a}{j} \frac{G_0}{G_{0a}} \frac{\cos(\theta - \psi_b)}{\sin(\psi_a - \psi_b)} - \frac{Z_{0b}H_b}{j} \frac{G_0}{G_{0b}} \frac{\cos(\theta - \psi_a)}{\sin(\psi_a - \psi_b)} \\ H &= H_a \frac{G_1}{G_{1a}} \frac{\sin(\psi - \psi_b)}{\sin(\psi_a - \psi_b)} + H_b \frac{G_1}{G_{1b}} \frac{\sin(\psi_a - \psi)}{\sin(\psi_a - \psi_b)} \end{aligned} \qquad [12]$$

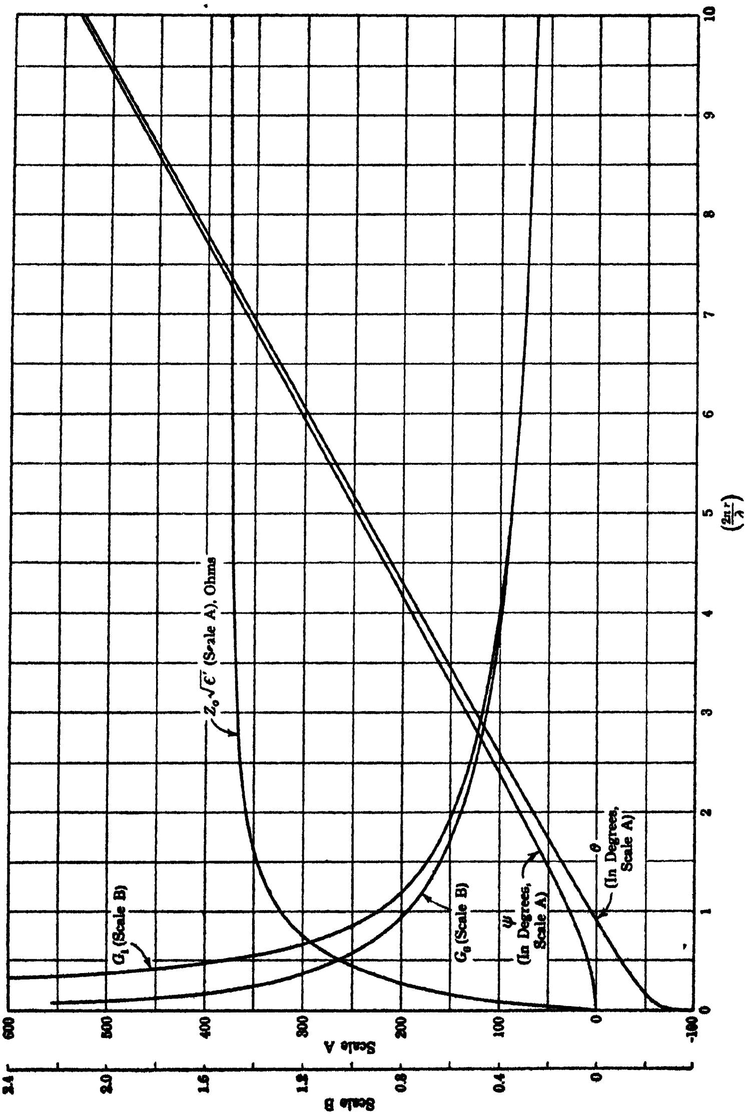


FIG. 9.08c. Radial transmission line quantities.

4. Input impedance $Z_i = \left.\frac{E_z}{H_\phi}\right|_i$ when load impedance $Z_L = \left.\frac{E_z}{H_\phi}\right|_L$ is given,

$$Z_i = Z_{0i}\left[\frac{Z_L \cos(\theta_i - \psi_L) + jZ_{0L}\sin(\theta_i - \theta_L)}{Z_{0L}\cos(\psi_i - \theta_L) + jZ_L \sin(\psi_i - \psi_L)}\right] \tag{13}$$

5. Input impedance $Z_i = \left.\frac{E_z}{H_\phi}\right|_i$ when output is shorted ($Z_L = 0$).

$$Z_i = jZ_{0i}\frac{\sin(\theta_i - \theta_L)}{\cos(\psi_i - \theta_L)} \tag{14}$$

6. Input impedance $Z_i = \left.\frac{E_z}{H_\phi}\right|_i$ when output is open circuited ($Z_L = \infty$).

$$Z_i = -jZ_{0i}\frac{\cos(\theta_i - \psi_L)}{\sin(\psi_i - \psi_L)} \tag{15}$$

Usually total current and voltage are desired before the problem is regarded as completely solved. They can be obtained from the field expressions. Total voltage, if a higher potential on the upper plate is considered positive,

$$V = -E_z d \tag{16}$$

Total current, if outward current in upper plate is considered as positive,

$$I = 2\pi r H_\phi \tag{17}$$

So that the relation between total impedance and those given above on a field basis, when the input end is at an inner radius,

$$Z_{\text{total}} = -\frac{d}{2\pi r}\left(\frac{E_z}{H_\phi}\right) \qquad r_i < r_L \tag{18}$$

If the input end is at an outer radius, the convention for positive current will be opposite to (17), so

$$Z_{\text{total}} = \frac{d}{2\pi r}\left(\frac{E_z}{H_\phi}\right) \qquad r_i > r_L \tag{19}$$

There are many higher order waves possible between the circular parallel conducting plates, having variations either with z, with ϕ, or both. If there are variations with z, the plates must be more than a half wave apart in order for the wave to propagate, just as was found for

the parallel plane transmission lines. The waves having variations with ϕ but not with z are more interesting since these will again propagate for any separation d between plates. In these, one part of the circumference may have fields in one direction whereas another part has fields in a reverse direction. These waves are of particular interest in sectoral electromagnetic horns.[2]

Problem 9.08(*a*). Set up the relations between fields that can exist in a sectoral horn, a region bounded by two parallel planes at $z = 0$, $z = d$ (in cylindrical coordinates) and two axial planes $\phi = 0$, $\phi = \phi_1$. Assume perfect conductors in the boundaries. Find the wave impedances for all wave types and the cut-off frequencies for the waves that exhibit cut-off phenomena.

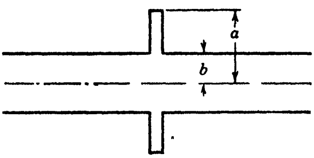


FIG. 9.08*d*. Circular wave guide with shorted radial line in series with cylinder wall.

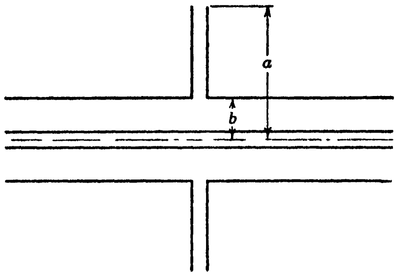


FIG. 9.08*e*. Coaxial line with open radial line in series with outer conductor.

Problem 9.08(*b*). For a TM_{01} wave in a circular wave guide it is desired to insert a blocking impedance for a given frequency. To do this, a section of shorted radial line (Fig. 9.08*d*) is inserted in the guide, its outer radius a chosen so that with the guide radius b given, the impedance looking into the radial line is infinite at the given frequency. Suppose that the radius b is 1.25 times greater than cut-off radius at this frequency for the TM_{01} wave and find the radius a.

Problem 9.08(*c*). It is sometimes required to break the outer conductor of a coaxial line for insulation purposes, without interrupting the r-f current flow. This may be accomplished by the radial line as shown (Fig. 9.08*e*) in which a is chosen so that with b and the operating wavelength specified, the radial line has zero input impedance seen from the line. Find the value of a assuming that end effects are negligible, and that

$$\frac{2\pi b}{\lambda} = 1$$

Problem 9.08(*d*). Find the voltage at the radius a in terms of the coaxial line's current flowing into the radial line at radius b (Fig. 9.08*e*).

9.09 Waves Guided by Conical Systems

The problem of waves guided by conical systems (Fig. 9.09) is important to a basic understanding of waves along dipole antennas and

[2] W. L. Barrow and L. J. Chu, *Proc. I.R.E.*, **27**, 51 (1939).

in certain classes of cavity resonators. In particular, one very important wave propagates along the cones with the velocity of light and has no field components in the radial direction, and so is analogous to the transmission line wave on cylindrical systems. This basic wave is symmetric about the axis of the guiding cones, so that if the two curl relations of Maxwell's equations are written in spherical coordinates with all ϕ variation eliminated, it is seen that there is one independent set containing E_θ and H_ϕ and E_r only:

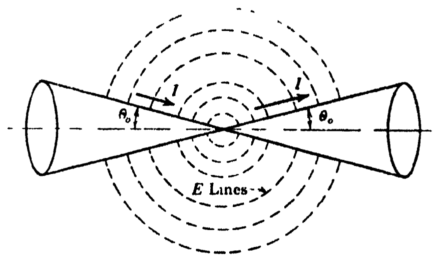


FIG. 9.09. Biconical guide.

$$\frac{1}{r}\frac{\partial(rE_\theta)}{\partial r} - \frac{1}{r}\frac{\partial E_r}{\partial \theta} + j\omega\mu_1 H_\phi = 0 \qquad [1]$$

$$\frac{1}{r \sin\theta}\left[\frac{\partial}{\partial\theta}(\sin\theta H_\phi)\right] - j\omega\epsilon_1 E_r = 0 \qquad [2]$$

$$-\frac{1}{r}\frac{\partial(rH_\phi)}{\partial r} - j\omega\epsilon_1 E_\theta = 0 \qquad [3]$$

Although we might proceed to a direct attack on these equations, it can be checked by substitution that the following solution does satisfy the three equations.

$$E_r = 0 \qquad [4]$$

$$rE_\theta = \frac{\eta_1}{\sin\theta}\left[Ae^{j(\omega t - k_1 r)} + Be^{j(\omega t + k_1 r)}\right] \qquad [5]$$

$$rH_\phi = \frac{1}{\sin\theta}\left[Ae^{j(\omega t - k_1 r)} - Be^{j(\omega t + k_1 r)}\right] \qquad [6]$$

$$k_1 = \omega\sqrt{\mu_1\epsilon_1} \qquad \eta_1 = \sqrt{\frac{\mu_1}{\epsilon_1}}$$

These equations show the now familiar propagation behavior, the first term representing a wave traveling radially outward with the velocity of light in the dielectric material surrounding the cones, the second term representing a radially inward traveling wave of the same velocity. The ratio of electric to magnetic field is given by $+\eta_1$ for the positively traveling wave, $-\eta_1$ for the negatively traveling wave. There is no field

component in the radial direction, which is the direction of propagation.

The above wave looks much like the ordinary transmission line waves of uniform cylindrical systems. This resemblance is stressed if we note that the E_θ corresponds to a voltage difference between the two cones,

$$V = -\int_{\theta_0}^{\pi-\theta_0} E_\theta r\, d\theta = -\eta_1 \int_{\theta_0}^{\pi-\theta_0} \frac{d\theta}{\sin\theta} [Ae^{j(\omega t - k_1 r)} + Be^{j(\omega t + k_1 r)}]$$

$$= 2\eta_1 \ln \cot \frac{\theta_0}{2} [Ae^{j(\omega t - k_1 r)} + Be^{j(\omega t + k_1 r)}] \qquad [7]$$

where the case treated is that of equal angle cones (Fig. 9.09). This is a voltage which is independent of r, except through the propagation term, $e^{\pm jk_1 r}$. Similarly the azimuthal magnetic field corresponds to a current flow in the cones,

$$I = 2\pi r H_\phi \sin\theta$$

$$= 2\pi[Ae^{j(\omega t - k_1 r)} - Be^{j(\omega t + k_1 r)}] \qquad [8]$$

This current is also independent of radius, except through the propagation term. A study of the sign relations shows that it is in opposite directions in the two cones at any given radius.

The ratio of voltage to current in a single outward traveling wave, a quantity which we call characteristic impedance in an ordinary transmission line, is obtained by setting $B = 0$ in (7) and (8):

$$Z_0 = \frac{\eta_1 \ln \cot \theta_0/2}{\pi} \qquad [9]$$

For a negatively traveling wave, the ratio of voltage to current is the negative of this quantity. This value of impedance is a constant, independent of radius, unlike those defined for a radial transmission line in Art. 9.08. We might have guessed this had we started from the familiar concept of Z_0 as $\sqrt{L/C}$ since inductance and capacitance between cones per unit radial length are independent of radius. This comes about since surface area increases proportionally to radius, and distance separating the cone, along the path of the electric field, also increases proportionally to radius.

So far as this wave is concerned, the system arising from two ideal coaxial conical conductors can be considered as a uniform transmission line. All the familiar formulas for input impedances and voltage and current along the line hold directly with Z_0 given by (9) and phase constant corresponding to velocity of light in the dielectric.

$$\beta = \frac{2\pi}{\lambda_1} = \omega\sqrt{\mu_1\epsilon_1} \qquad [10]$$

If the conducting cones have resistance there is a departure from uniformity due to this resistance term, but this is usually not serious in any practical cases where such conical systems are used.

Of course a large number of higher order waves may exist in this conical system and in other similar systems. These will in general have field components in the radial direction and will not propagate at the velocity of light. We shall consider such general wave types for spherical coordinates later.

Problem 9.09(*a*). It has been seen that along cones, cylinders, planes, etc., a principal wave can exist in which, at least for perfect conductors, it is possible to analyze the problem correctly by dealing with distributed L's and C's per unit length, these distributed constants being computed from static field distributions. It is not always true that the electric and magnetic field lines over the cross section of the wave, for the principal wave, will be as in the static case. This does not mean that the distributed constant technique fails for such lines, but it does mean that it is no longer exact to use L and C as computed from the static field equations. Illustrate the above statement by considering waves propagating symmetrically between concentric spheres in the θ direction ($\partial/\partial\phi = 0$). Show that no wave can exist containing only E_r and H_ϕ; E_θ must also be present. Show also that if the distance between spheres is small compared with wavelength, the presence of E_θ has a negligible effect on the wave distribution and distributed L and C (computed from statics) may be used for good approximate results.

Problem 9.09(*b*). Derive the basic characteristics of the principal waves on a transmission line consisting of two coaxial, common-apex cones of unequal angles.

9.10 Wave Guides of Special Cross Section

There is an interesting point of view that is especially useful in thinking about wave guides of certain special shapes. Recall first that it was shown in Chapter 8 that the properties of an ideal guide (propagation constant, wave impedance, etc.) are determined once the cut-off frequency is determined, and have the same forms for all shapes of guides. Thus the analysis of the ideal guide requires only the determination of the cut-off frequency. This has been done in previous examples by the solution of a differential equation subject to boundary conditions. It may be recognized, however, that the cut-off frequency for a given mode corresponds to resonance for waves propagating only transversely in the given cross section, according to that mode. That is, it is necessary only to find the resonant frequency of the two-dimensional problem defined by the guide boundary, and this will be the cut-off frequency of the guide. (Actually, there will, of course, be an infinite number of possible resonances, each corresponding to a cut-off for a given wave type.)

The use of resonance in the two-dimensional problem does give cut-off frequency since there are no axial variations at cut-off; all energy does

propagate back and forth in the transverse plane. Thus, for example, the TE_{10} wave in a rectangular wave guide had a cut-off frequency equal to the resonant frequency for a plane wave propagating only in the x direction across the guide, thus corresponding to a half wavelength in the x direction. The TM_{01} wave in a circular guide had cut-off frequency equal to the resonant frequency for waves propagating radially with E_z and H_ϕ. These and other similar examples may be verified by an inspection of previous results.

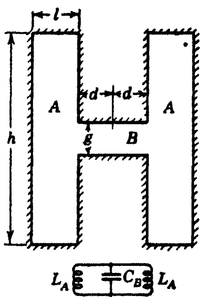


FIG. 9.10. Cross section of special wave guide and approximate equivalent circuit for cut-off calculation.

The point of view is useful when guides are being studied of such shape that resonance may be obtained simply by approximate methods. Thus in Fig. 9.10, the shape of guide is such that region B, for the simplest mode, can be considered essentially as a capacitor of capacity (for a unit length)

$$C_B = \frac{\epsilon_1(2d)}{g} \tag{1}$$

The regions A can be considered essentially as inductances, of value (for a unit length)

$$L_A = \mu_1 lh \tag{2}$$

The approximate equivalent circuit for resonance calculations is thus as drawn in Fig. 9.10, and resonant frequency, which is cut-off frequency for this mode, is

$$f_c \cong \frac{1}{2\pi\sqrt{C_B L_A/2}} = \frac{1}{2\pi}\sqrt{\frac{g}{\mu_1\epsilon_1 lhd}} \tag{3}$$

$$\lambda_c = \frac{1}{f_c\sqrt{\mu_1\epsilon_1}} \cong 2\pi\sqrt{\frac{lhd}{g}} \tag{4}$$

Note in this example the *theoretical* possibility of making f_c as low as is desired for a given cross section, simply by decreasing g indefinitely. Thus it is not necessary that transverse dimensions be comparable to wavelength, for they may be "foreshortened" as in this example in a manner quite analogous to the capacity foreshortening of resonant lines.

More accurate calculations of this and similar examples would require procedures such as those to be outlined in Art. 9.15. Also, since this procedure is now reduced to one of calculating resonances, many of the methods to follow in Chapter 10 may be applied.

SPECIAL PROBLEMS IN GUIDED WAVE APPLICATIONS

9.11 Excitation and Reception of Waves in Guides

The problems of exciting waves in wave guides and of absorbing their energy in a receiver are extremely difficult to analyze if exact quantitative analysis is desired. The qualitative picture is not difficult. In order to excite any particular desired wave, one should study the wave pattern, and then use any of the following methods.

1. Introduce the excitation in a probe or antenna which is placed at the point of maximum electric field, oriented in the direction of the electric field.
2. Introduce the excitation through a loop which is placed at the point of maximum magnetic field, the plane of the loop being normal to the magnetic field.
3. Introduce currents from transmission lines or other sources in such a manner that the desired current directions in the guide walls are forcibly excited. (Of course it is true that since currents and fields are directly related, any scheme based on exciting currents in the walls may, if preferred, be looked upon as a scheme of exciting fields in the space, but the viewpoint from currents is often more direct.)
4. For higher order waves combine as many of the exciting sources as are required, with proper phasings.

Since any of the above exciting methods are in the nature of concentrated sources, they will not in general excite purely one wave, but all waves which have field components in a favorable direction for the particular exciting source. From another point of view, we see that one wave alone will not suffice to satisfy the boundary conditions of the guide complicated by the exciting source, so that many higher order waves must be added for this purpose. If the guide is large enough, several of these waves will then proceed to propagate. Most often, however, only one of the excited waves is above cut-off. This will propagate down the guide, and (if absorbed somewhere) will represent a resistive load on the source, comparable to the radiation resistance of antennas which we shall encounter further in Chapter 11. The higher order waves which are excited, if all below cut-off, will be localized in the neighborhood of the source and will represent purely reactive loads on the source. For practical application, it is then necessary to add, in the line which feeds the probe or loop or other exciting means, an arrangement for matching to the load which has a real part representing the propagating wave and an imaginary part representing the localized reactive waves.

The receiving problem is the reverse of the exciting problem, and in

general any method which works well for exciting will also work well for receiving.

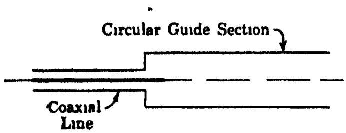


FIG. 9.11a. Antenna in end of circular guide for excitation of TM_{01} wave.

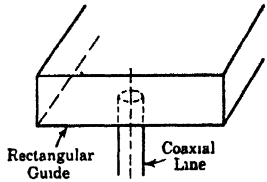


FIG. 9.11b. Antenna in bottom of rectangular guide for excitation of the TE_{10} wave.

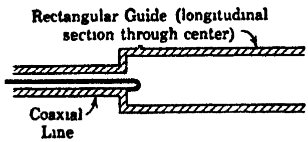


FIG. 9.11c. Loop in end of rectangular guide for excitation of TE_{10} wave.

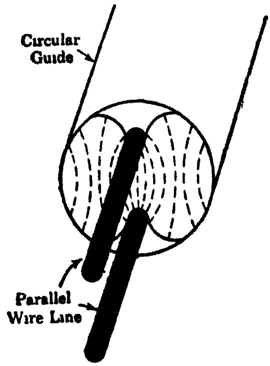


FIG. 9.11d. Parallel wire line for excitation of TE_{11} wave in circular guide.

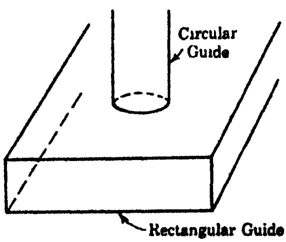


FIG. 9.11e. Junction between circular guide (TM_{01} wave) and rectangular guide (TE_{10} wave.)

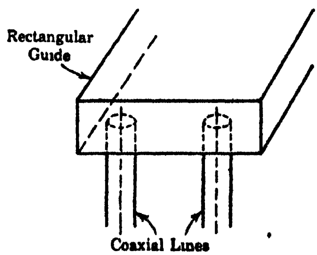


FIG. 9.11f. Excitation of the TE_{20} wave in rectangular guide by two oppositely phased antennas.

Some examples of the several excitation methods listed in 1 to 4 are shown in the Figs. 9.11a to 9.11f. In Fig. 9.11a an antenna is used to excite a TM_{01} wave in a circular guide. In Fig. 9.11b a similar antenna

is used to excite a TE_{10} wave in a rectangular guide. Note that one end of the guide is closed to obtain transmission in one direction only. The position of this closed end may be utilized as one variable in the matching process. In Fig. 9.11*c*, a TE_{10} wave in a rectangular guide is excited by a loop. In Fig. 9.11*d*, a TE_{11} wave in a circular guide is excited by the currents of a two-wire transmission line. Similarly, in Fig. 9.11*e*, the TM_{01} wave in the circular guide is excited by a TE_{10} wave in a rectangular guide, and a study of the patterns and reflections at the closed end shows that currents in the walls are proper here for excitation. Finally in Fig. 9.11*f*, a TE_{20} wave in a rectangular guide is excited by two antennas, properly phased. Further discussion, with experimental verification is presented by Southworth.[3]

Problem 9.11. Draw the field and current patterns in the lines and guides of Figs. 9.11*a* to 9.11*f*, and explain the coupling mechanism in each of these figures. Discuss the impedance matching problem for these cases.

9.12 Transmission Line Techniques Applied to Wave Guides

The transmission line analogy for propagation of uniform plane waves was developed extensively in previous chapters (Arts. 7.07ff.). This is a rigorous analogy and is useful for two major reasons. First, the propagation characteristics of uniform plane waves may be analyzed from expressions already developed and well known for transmission lines. More important, the well-known transmission line techniques (quarter-wave matching sections, methods of termination, etc.) may be applied directly to plane waves by means of the analogy.

The above analogy was used as follows. If plane discontinuities existed, a direction was chosen normal to those discontinuities, and the phase velocity calculated in that direction (even though the wave may have been looked at previously as propagating in some other direction). The ratio of the component of electric field transverse to the selected direction (and therefore parallel to the plane discontinuity) to the transverse magnetic field component was defined as a wave impedance. This was the quantity used in place of actual impedance in the transmission line equations. The transverse component of electric field was then analogous to voltage along a transmission line; the transverse component of magnetic field was analogous to current. It should be evident that the same analogies may be applied directly to all the guided waves studied in this chapter, so long as the guides do not possess discontinuities. Discontinuities can be allowed without interfering with the accuracy of the method only if they take the form of a variation in the dielectric of the guide with the discontinuity boundaries always normal to the

[3] G. C. Southworth, *Proc. I.R.E.*, **25**, 807–822, July, 1937.

direction of propagation and extending over the entire cross section; the guide cannot be permitted to change either its shape, section, or direction at this point for the wave impedance concept to be directly applicable. Other types of discontinuities will require a consideration of total quantities.

Characteristic wave impedances have been defined for each wave type. We shall now list some of the resulting transmission line techniques which may be directly applied to wave guides by use of the transmission line analogy with wave impedances.

For convenience, the wavelength along the direction of the guide will be labeled λ_g where

$$\lambda_g = \frac{\lambda_1}{\sqrt{1 - (f_c/f)^2}} = \frac{1}{f\sqrt{\mu_1\epsilon_1}\sqrt{1 - (f_c/f)^2}} \qquad [1]$$

The phase constant for all waves,

$$\beta = \frac{2\pi}{\lambda_g} = \omega\sqrt{\mu_1\epsilon_1}\sqrt{1 - (f_c/f)^2} \qquad [2]$$

The wave impedance

Transverse electromagnetic waves $\quad Z_{TEM} = \eta_1 = \sqrt{\dfrac{\mu_1}{\epsilon_1}} \qquad [3]$

Transverse magnetic (E) waves $\quad Z_{TM} = \eta_1\sqrt{1 - (f_c/f)^2} \qquad [4]$

Transverse electric (H) waves $\quad Z_{TE} = \dfrac{\eta_1}{\sqrt{1 - (f_c/f)^2}} \qquad [5]$

Notice in particular that the wavelength along the guide for either TM or TE waves is always longer than the corresponding wavelength for transverse electromagnetic waves, i.e., uniform plane waves and transmission line waves. The wave impedance of TM waves is always less than the intrinsic impedance of the medium, η_1; the wave impedance for TE waves is always greater than η_1.

Short-Circuited Guide. A wave guide may be considered as truly short-circuited if a conducting plate is placed across the entire section of the guide so that the transverse component of electric field is reduced to zero over all that section. This corresponds to a shorted transmission line, so that at once we may draw the forms of the resulting standing wave pattern (Fig. 9.12a). Transverse electric field is zero at the conducting plate and at multiples of $\lambda_g/2$ in front of it. It is a maximum at odd multiples of $\lambda_g/4$ in front of the plate. Transverse magnetic field is a maximum at the plate and has other maxima at $n\lambda_g/2$; minima at

$(2n + 1)\lambda_g/4$ before the plate. Other phase relations show that E_z for TM waves has the same axial distribution pattern as the magnetic field, and H_z for TE waves has the same axial distribution pattern as the electric field.

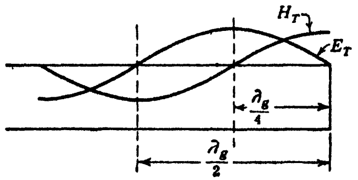


FIG. 9.12a. Standing waves of transverse field components in shorted guide.

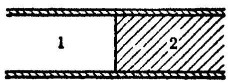


FIG. 9.12b. Guide with dielectric discontinuity.

Guide with Dielectric Discontinuity. If there is a discontinuity from one dielectric to another in a guide (Fig. 9.12b), the amount of reflection into the first region and the transmission to the second region may be determined from the mismatch in impedances Z_1 and Z_2. The expressions are:

$$\frac{E'_{T_1}}{E_{T_1}} = \frac{K-1}{K+1} = -\frac{H'_{T_1}}{H_{T_1}}$$

$$\frac{E_{T_2}}{E_{T_1}} = \frac{2K}{K+1} = \frac{KH_{T_2}}{H_{T_1}} \qquad [6]$$

$$K = \frac{Z_2}{Z_1}$$

The region (1) then has both a standing wave and a traveling wave. The other standard expressions for input impedance, and voltage and current along the line, from Chapter 1, may be applied to calculation of input impedance on a field basis and of values of electric and magnetic fields along the guide.

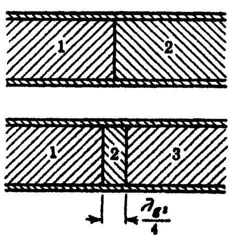


FIG. 9.12c. Insertion of matching section in a guide.

Quarter-Wave Matching Sections. It is of course possible to match between one section of a guide and another section with different dielectric constant for any of the wave types at any single frequency. This is accomplished by the technique of quarter-wave matching sections developed for transmission lines in Prob. 1.21(b) and for plane waves in Art. 7.09. Thus in Fig. 9.12c it is possible to

match between the regions 1 and 3, if a region 2 is introduced, a quarter wavelength long (measured at the phase velocity in that region) and having an impedance the geometric mean of those on the two sides. Note that in calculating these impedances, the different cut-off frequencies for the three sections must be taken into account in (4) or (5).

This matching may be used, for instance, in a case where it is desired to absorb power in section (3), which may be filled with water or some other material with a small but finite conductivity. A quarter-wave section of a proper material (certain special glasses, for example) may then be used to match this section to the portion of the guide with air dielectric.

Elimination of Reflections from Dielectric Slabs. If dielectric slabs must be placed in an otherwise uniform guide (for example, because a section must be evacuated), these may be designed in certain ways so that they cause no reflections, just as may insulators in transmission lines. The simplest arrangement is to make the dielectric slabs a half wavelength in thickness ($\lambda_g/2$ for the material of the slab). The impedance at the front is then exactly the impedance of the guide following the slab. From another point of view, the reflections from the front and back surfaces exactly cancel under these conditions.

The above method of eliminating reflections requires that the dielectric slab be a half wavelength in thickness, measured in the material of that slab. For certain applications it may be undesirable to use slabs of that thickness. For slabs of any thickness, reflections may be eliminated by cancelling the reflected wave from one slab by that from another placed a proper distance from it. For slabs of thickness small compared with wavelength, or of a material with properties not too greatly different from that of region 1, this spacing is such that the total phase angle corresponding to the length of guide between insulators and one insulator is very nearly 90°.

Termination of Wave Guides. Another important technique of transmission lines is the termination of a line by means of a proper resistor to eliminate the reflected wave. All energy is completely absorbed according to the simple line theory if this resistor is equal to the characteristic wave impedance of the line. If the line must be closed at the end, the terminating resistor may be placed a quarter wavelength from the shorted end, since for perfect conductors the shorted quarter-wave line represents an infinite impedance in parallel with the resistance. Similarly, a wave guide may be terminated by a conducting sheet having a resistance per unit square equal to the characteristic wave impedance of the wave type to be matched. This sheet is placed a quarter wave-

length from the shorted end (Fig. 9.12d).

$$d_3 = \frac{\lambda_{g3}}{4}$$

$$\frac{1}{\sigma_2 d_2} = Z_1$$

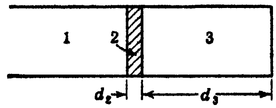


FIG. 9.12d. Conducting film for terminating a guide.

Notice that the conducting film must be made of some material of relatively low conductivity if its thickness is not to be absurdly small. That is, for a material like copper, d_2 would be only of the order of 10^{-10} meter.

Problem 9.12. Show, for a TM wave in any shape of guide passing from one dielectric material to another, that at one frequency the change in cut-off factor may cancel the change in η, and the wave may pass between the two media without reflection, even though no intervening matching section is present. Identify this condition with the case of incidence at polarizing angle in Art. 7.14. Determine the requirement for a similar situation with TE waves and show why it is not practical to obtain this.

9.13 Waves Below Cut-Off

The higher order waves which may exist in coaxial lines and all waves which may exist in hollow pipe wave guides are characterized by cut-off frequencies. If the waves are to be used for propagating energy, we are of course interested only in the behavior above cut-off. However, the behavior of these waves, perhaps we should say "imaginary waves," below cut-off is important in at least two practical cases:

1. Application to wave guide attenuators.
2. Effects at discontinuities in transmission systems.

The attenuation properties of these waves below cut-off have been developed in the previous analyses. It has been found that below the cut-off frequency there is an attenuation only and no phase shift in an ideal guide. The characteristic wave impedance is a purely imaginary quantity, again emphasizing the fact that no energy can propagate down the guide. This is not a dissipative attenuation as is that due to resistance and conductance in transmission systems with propagating waves. It is a purely reactive attenuation, analogous to that in a filter section made of reactive elements, when this is in the cut-off region. The energy is not lost but is reflected back to the source so that the guide acts as a pure reactance to the source.

The expression for attenuation below cut-off in an ideal guide, Eq. 8.18(4), may be written

$$\gamma = \alpha = k_c\sqrt{1 - \left(\frac{f}{f_c}\right)^2} = \frac{2\pi}{\lambda_c}\sqrt{1 - \left(\frac{f}{f_c}\right)^2} \qquad [1]$$

As f is decreased below f_c, α increases from a value of 0 approaching the constant value

$$\alpha = \frac{2\pi}{\lambda_c} \tag{2}$$

when $(f/f_c)^2 \ll 1$. This is an important point in the use of wave guide attenuators, since it shows that the amount of this attenuation is substantially independent of frequency if the operating frequency is very far below the cut-off frequency. In addition, the amount of this attenuation is determined only by the cut-off wavelength of the guide, which is in general proportional to the transverse size of the guide, so that the value of α may be made almost as large as one pleases by selecting a low cut-off wavelength (small pipe size). Since (1) holds for any wave in any shape of guide, it follows that choices of wave type and guide shape cannot influence the attenuation constant except in so far as they fix the cut-off wavelength λ_c.

Note that if a wave guide attenuator is designed with $(f/f_c) \ll 1$ so that attenuation is independent of frequency, attenuation must necessarily be very great in a wavelength since α will be much greater than the free space phase constant,

$$\frac{\alpha}{k_1} = \frac{2\pi/\lambda_c}{2\pi/\lambda_1} = \frac{\lambda_1}{\lambda_c} \gg 1$$

Now let us look for a moment at the relations among the fields of both transverse magnetic and transverse electric waves below cut-off. If $\gamma = \alpha$ as given by (1) is substituted in the expressions for field components of transverse magnetic waves, Eq. 8.11(3),

$$\begin{aligned} H_x &= \frac{j}{\eta_1}\left(\frac{f}{f_c}\right)\frac{1}{k_c}\frac{\partial E_z}{\partial y} & E_x &= -\sqrt{1-(f/f_c)^2}\,\frac{1}{k_c}\frac{\partial E_z}{\partial x} \\ H_y &= -\frac{j}{\eta_1}\left(\frac{f}{f_c}\right)\frac{1}{k_c}\frac{\partial E_z}{\partial y} & E_y &= -\sqrt{1-(f/f_c)^2}\,\frac{1}{k_c}\frac{\partial E_z}{\partial x} \end{aligned} \tag{3}$$

For a given distribution of E_z across the guide section, which is determined once the guide shape and size and the wave type are determined, it is evident from the relations (3) that as frequency decreases, $f/f_c \to 0$, the components of magnetic field approach zero whereas the transverse components of electric field approach a constant value. We draw the conclusion that only electric fields are of importance in transverse magnetic or E waves far below cut-off. Similarly, only magnetic fields are of importance in transverse electric or H waves far below cut-off.

Suppose a TM wave is excited by some source in a wave guide, extend-

ing down the guide a certain distance to a suitable receiver. If the frequency is far enough below cut-off so that $(f/f_c)^2$ is negligible compared with unity, the entire problem may be looked upon as one of electric coupling between the source and the receiver, calculated by D-C or low-frequency methods (of course, taking into account the presence of the guide as a shield). Similarly, a *TE* wave between a source and receiver in a guide far below cut-off may be looked upon as a problem of ordinary magnetic coupling between the source and receiver (of course, taking into account the presence of the guide as a shield). That this is true may be seen from the following. If the waves are far below cut-off, the dimensions of the guide must be small compared with wavelength. Fields will then attenuate to a negligible amount in a distance small compared with wavelength. For any such region small compared with wavelength, the wave equation will reduce to Laplace's equation so that low-frequency analyses neglecting any tendency toward wave propagation are applicable. Only when dimensions of the guide become large enough compared with wavelength so that $(f/f_c)^2$ is comparable to unity, must the tendency toward propagation be considered. That is, the effects of magnetic fields must be considered in *TM* waves, and the effects of electric fields in *TE* waves.

9.14 Waves in the Vicinity of Cut-Off

The cut-off frequency for a guide with lossless conductors and dielectric would be a definite frequency at which attenuation would pass from a finite value or all lower frequencies to a value of zero for all higher frequencies. The phase constant β would be zero for all frequencies below cut-off and finite for all frequencies above. Such ideal curves are sketched in the heavy curves of Fig. 9.14. It has already been shown that imperfect conductors or dielectric introduce a finite attenuation at frequencies above cut-off, and may also change the phase constant somewhat. Similarly, imperfect conductors or dielectrics will act to produce a small but finite phase shift below cut-off and a certain correction to attenuation. The cut-off frequency under such conditions no longer represents a sharp transition but a more gradual change from one region to the other. This is indicated by the dotted curves of Fig. 9.14.

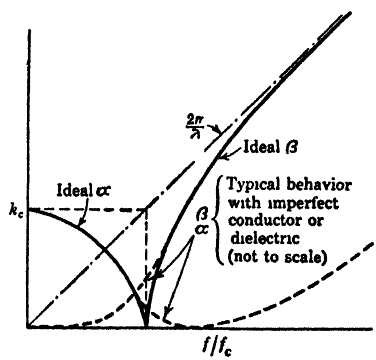


FIG. 9.14. Modification of propagation characteristics due to losses.

It should be emphasized again that many of the approximate formulas developed so far are extremely inaccurate for frequencies very close to cut-off. For example, the formulas for attenuation due to imperfect conductors or dielectrics, Eqs. 8.18(10) and 8.18(11), will show an incorrect value of $\alpha = \infty$ at $f = f_c$. Similarly, phase velocity will appear incorrectly to be infinite at $f = f_c$ by Eq. 8.18(5).

9.15 Discontinuities in Lines and Guides

Higher order waves are also important at discontinuities in transmission systems. As an example of their occurrence at such discontinuities, consider the step in the parallel plane transmission line of Fig. 9.15*a*. In a transmission line wave (principal wave) between planes there are E_y and H_x only, and no variations with y. The perfect conductor portion from (2) to (3) requires that $E_y = 0$ here. If there were only principal waves, E_y would then have to be zero everywhere at $z = 0$ because of the lack of variations with y in the principal wave. There could then be no energy passing into the second line A regardless of its termination since the Poynting vector would then also be zero across the entire plane $z = 0$. Physical reasoning shows that the above situation does not occur generally but only in such special cases as when line A is shorted a half wave from the discontinuity. The difficulty is met by the higher order waves which are excited at the discontinuity, so that E_y in the principal wave is not zero at $z = 0$, but *total* E_y (sum of principal and higher order components) is zero from (2) to (3) but not from (1) to (2). For the example of Fig. 9.15*a* the higher order waves excited are *TM* waves since E_y, E_z, and H_x alone are required in the fringing fields. For spacings between planes not comparable to wavelength, these waves are far below cut-off so that their fields are localized in the region of the discontinuity. They may consequently be called *local* waves.

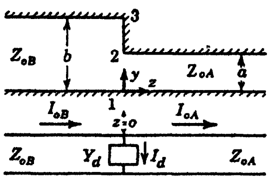


FIG. 9.15*a*. Step discontinuity in parallel plane transmission line and exact equivalent circuit.

The example considered is one of those which may be solved mathematically, although the solution will not be detailed here. The method consists in setting down the series of principal wave and higher order *TM* wave solutions in each region of the line, A and B. The amounts of the higher order waves are determined by matching the tangential field components across the boundary: total E_y in the B region must equal E_y in the A region from (1) to (2), and must equal zero from (2) to (3); H_x in the A region must equal H_x in the B region from (1) to (2). The

resulting relations between the series may be handled with reasonable ease when put in a form such that series tabulated by Hahn[4] may be used.

Perhaps the most important information which comes from such an exact analysis is the fact that a transmission line equivalent circuit can be drawn with line A joined to line B and a lumped admittance placed at $z = 0$ to account for the effects of the local waves. Thus if current $I(z)$ is found in either of the planes at any value of z, it has a contribution $I_0(z)$ from the principal wave and a contribution $I'(z)$ from all local waves.

$$I(z) = I_0(z) + I'(z) \qquad [1]$$

Now *total* current must be continuous at the discontinuity $z = 0$, but current in the principal wave need not be since the difference in principal wave currents may be made up by the local wave currents.

$$I_{0A}(0) + I'_A(0) = I_{0B}(0) + I'_B(0)$$

or

$$I_{0B}(0) - I_{0A}(0) = I'_A(0) - I'_B(0) \qquad [2]$$

However, total voltage in the line as defined from $-\int \bar{E} \cdot d\bar{l}$ between planes is only that in the principal wave, since a study of the local waves shows that their contribution is zero.

$$V(z) = V_0(z)$$

Continuity of total voltage across the discontinuity $z = 0$ then requires continuity of voltage in the principal wave.

$$V_{0A}(0) = V_{0B}(0) = V_0(0) \qquad [3]$$

Now if an equivalent circuit is drawn *for the principal wave only*, its continuity of voltage but discontinuity of current may be accounted for by a lumped *discontinuity admittance* at $z = 0$, the current through this admittance being

$$I_{0B}(0) - I_{0A}(0) = I_d = Y_d V_0(0)$$

Or, from (2)

$$Y_d = \frac{I'_A(0) - I'_B(0)}{V_0(0)} \qquad [4]$$

[4] W. C. Hahn, "A New Method for the Calculation of Cavity Resonators," *Journ. Appl. Phys.*, **12**, 62–68 (January, 1941).

The complete analysis[5] reveals that when local wave values are substituted in (4), numerical values of Y_d may be calculated which are independent of terminations so long as these are far enough removed from

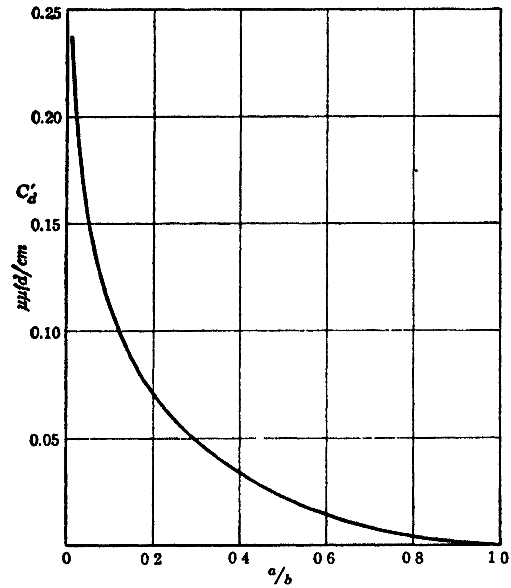


FIG. 9.15*b*. Curve of discontinuity capacitance for Fig. 9.15*a*.

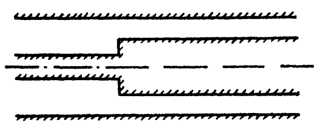


FIG. 9.15*c*. Typical discontinuity in coaxial line.

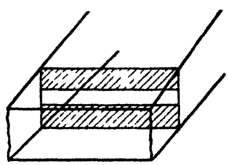


FIG. 9.15*d*. Capacitive diaphragm in rectangular guide.

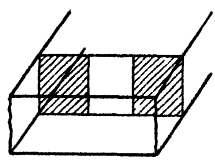


FIG. 9.15*e*. Inductive diaphragm in rectangular guide.

the discontinuity not to couple to the local wave fields. For Fig. 9.15*a*, with dimensions small compared with wavelength, this admittance turns out to be a pure capacitance, values of which are plotted versus step ratio a/b in Fig. 9.15*b*, in micromicrofarads per centimeter width of the plane. If multiplied by proper circumference, these values may be used to give the approximate discontinuity capacitance for corresponding steps in coaxial lines (Fig. 9.15*c*); they may also be extended to give numerical

[5] "Equivalent Circuits for Discontinuities in Transmission Lines," J. R. Whinnery and H. W. Jamieson. *Proc. I.R.E.*, **32**, 98–114 (February, 1944).

results for capacitive steps or diaphragms in rectangular wave guides with TE_{10} waves. (Fig. 9.15*d*.)

In all the above examples the susceptance to be lumped at the discontinuity is capacitive. In a case such as the diaphragm extending from the sides of a rectangular guide for use with the TE_{10} wave (Fig. 9.15*e*), the local waves excited are *TE* waves. By the reasoning of Art. 9.13 the energy in these should be magnetic so that such a discontinuity is inductive rather than capacitive.

Problem 9.15(*a*). Determine the form of the proper local waves in the example of Fig. 9.15*a*. Show that voltage between planes, $-\int \bar{E}\cdot\bar{dl}$, is zero for each of these.

Problem 9.15(*b*). Imagine a parallel plane transmission line with two steps such as the one in Fig. 9.15*a*. The first is from spacing b to spacing a; the second is removed from the first by a half wavelength and is from spacing a back to b. The line to the right of b is perfectly terminated by its characteristic impedance, Z_{0B}. If it were not for the discontinuity capacitances, the line to the left of the first discontinuity would also be perfectly terminated. Calculate reflection coefficient in this line taking into account the discontinuity capacitances from Fig. 9.15*b*. Take $a = 1$ cm, $b = 2$ cm, $\lambda = 12$ cm.

Problem 9.15(*c*). Using Fig. 9.15*b*, calculate an approximate discontinuity capacitance for the coaxial line of Fig. 9.15*c*. Take $r_1 = 0.5$ cm, $r_2 = 1$ cm, $r_3 = 1.2$ cm.

10

RESONANT CAVITIES

10.01 Introduction

At extremely high frequencies (wavelengths, say, below 1 meter) ordinary lumped circuit elements are hardly suitable for practical use. As was seen in Chapter 5, a conventional circuit with dimensions comparable to wavelength may lose energy by radiation. In Chapter 6 it was found that resistance of ordinary wire circuits may become high because of skin effect behavior. Both of these phenomena give rise to definite modifications in elements that are to serve as efficient circuits for ultra-high frequencies. It is immediately suggested that the circuit region should be shielded, completely surrounded by a good conductor, to prevent radiation. It is also suggested that the current paths be made with as large area as possible. The result is a hollow conducting box with the electromagnetic energy confined on the inside.[1] The conducting walls act effectively as perfect shields so that this inner region is perfectly shielded from the outside, and no radiation is possible. Since the inner walls of the box serve as current paths, the desired large area for current flow is provided and losses are extremely small. The resulting element is known as a cavity resonator.

In this chapter we shall study electromagnetic waves in regions closed by conductors, with particular application to such cavity resonators. It will first be observed that such high-frequency elements might be arrived at by extension of conventional transmission line and circuit ideas, and that circuit concepts, such as inductance, capacitance, and Q, may be used to great advantage in most cavity resonators. Exact analyses will be made of certain of the simpler shapes of cavity resonators, and at least approximate analyses will be made of some of the more complex shapes of such resonators. All mathematical analyses will be based on the solution of Maxwell's equations subject to the boundary conditions, and in general will follow directly from the results of the last several chapters on propagating waves, since the waves inside the conducting boxes may be considered as standing wave patterns arising from reflections of the appropriate traveling waves from the walls of the enclosure.

[1] W. W. Hansen, *Journ. Appl. Phys.*, 9, 654–663 (October, 1938).

SOME SIMPLE CAVITIES AND LUMPED CIRCUIT ANALOGIES

10.02 Elemental Analogies Leading to the Concept of Cavity Resonators

Before the solution of the wave equation inside regions closed by conductors is attempted, there are several physical analogies that should make the concept of such wave regions more meaningful, particularly in their function as " circuits " at ultra-high frequencies.

For the first analogy, let us consider something which is not ordinarily thought of as a cavity resonator, but which certainly may be. This is a section of coaxial transmission line shorted at both ends. From the transmission line analysis of Chapter 1, it is known that such a shorted line may support a standing wave of frequency such that the length of line is exactly a half wave. The line may be thought of as resonant at that frequency, since the standing wave pattern set up has constant total energy in that section of line, that energy oscillating between the electric and magnetic fields of the line. Thus, as in Fig. 10.02a, the standing wave of voltage has a zero at each end and a maximum at the center. The standing wave of current is 90° out of time phase with the voltage wave, and has maxima at the two ends, a zero at the center. These waves may exist inside this completely enclosed region without interference from, or radiation to, the outside. The shielding is complete if conductors are perfect, and practically so for any practical conductors at ultra-high frequencies. This viewpoint is verified by the previous analyses (Chapter 6) of skin effect phenomena, where it was found that depth of penetration at high frequencies is so small (of the order of 10^{-4} inch for copper at 3000 mc/sec) that almost any practical thickness acts essentially as an infinite thickness. Fields applied on the inside of a conducting wall die out to a completely negligible value at the outside of the conductor.

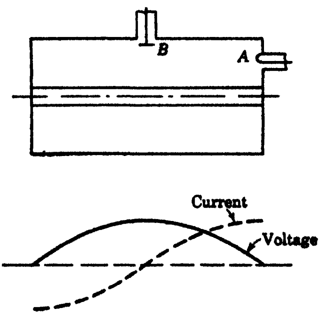


FIG. 10.02a. Resonant coaxial system and standing waves of voltage and current.

Since the inside of the region is completely shielded from the outside, it will be necessary to excite the waves by some source, such as the small loop A (Fig. 10.02a), designed to excite the magnetic field of the line at its maximum value, or the small probe B introduced at the maximum of electric field. If one of these means is used to stimulate the line exactly at its resonant frequency, the oscillations may build up to a large

value. In the steady state limit, the exciting source need supply only the relatively small amount of energy lost to the finite conductivity of the walls, the relatively large stored energy being essentially constant and passing back and forth between electric and magnetic fields. If the source excites the line at a frequency somewhat off resonance, the energies in electric and magnetic fields do not balance. Some extra energy must be supplied over one part of the cycle which is given back to the source over another part of the cycle, and the line acts as a reactive load on the exciting source in addition to its small loss component. The similarity to ordinary tuned circuit operation is evident, and it seems likely that many of the same considerations concerning effect of losses on band width, expressed in terms of a Q, will hold, at least qualitatively.

The above simple example requires essentially only a knowledge of transmission line theory, yet it holds all the fundamental characteristics of cavity resonators, and differs from others only in the types of waves that are utilized.

Since closed resonant cavities take the place of lumped L-C circuits at high frequencies, we shall see as a next example how a closed cavity might be considered as the logical evolution of such a circuit as it is extended to these frequencies. If a parallel resonant circuit with lumped L and C, such as that of Fig. 10.02*b*, is to be extended to high frequencies, a decrease must be made in the magnitudes of C and L. Capacitance may be decreased simply by moving the plates of the condenser farther apart. To decrease inductance, fewer and fewer turns might be used in the inductance until this has degenerated to a single straight wire. Next, to eliminate stray lead inductances, this might be moved to the condenser plates and connected directly between them at the edges. The final step suggested is the paralleling of many of these single-wire inductances about the outside of the plates, until in the limit the two plates are connected by a solid conducting wall. We are now left with a hollow cylindrical conducting box, completely enclosed, or in other words, another example of a cavity resonator.

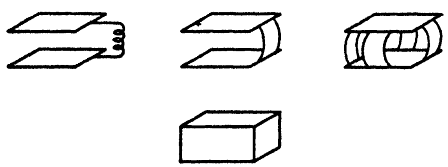


FIG. 10.02*b*. Evolution from resonant circuit with lumped elements to a closed cavity.

The above example is, of course, not exactly rigorous. It is significant in demonstrating a logical evolution from lumped circuit ideas to the concept of cavity resonators, but if only a knowledge of lumped circuits without any background in wave phenomena were available, there would be reason to doubt that the system arrived at in the limit

would even work. Certainly there is a point in the evolution where one realizes that the fields of the capacity and the inductances are becoming intimately related, and at best it is a problem with distributed rather than lumped constants with perhaps mutual impedances also present. It would appear safe to conclude that the condenser plates have actually been shorted in the limit, so that if any voltage can exist between them, it can only exist at the center and must form a standing wave pattern inside the box, falling to zero at the shorting walls, and so requiring that the box have a diameter at least comparable to wavelength. Here it may be protested that the side walls have been imagined to act as an inductance. How can there be always zero voltage across these walls then, since there is a voltage drop across an inductance whose current is changing? The answer involves recognizing that we are speaking of total voltage, and that total voltage across any inductance made of a perfect conductor must be zero, the applied voltage being exactly balanced by that induced from the changing magnetic fields of the inductance. But these are all tentative and preliminary pictures. We will not try to squeeze further conclusions from the present analogy, since it is realized that the wave picture is in reality the correct one and will determine whether any particular result or physical picture is legitimate. However, it will prove useful to recall this analogy from time to time in seeking circuit ideas that may be employed in discussing resonator behavior.

A third picture of the electromagnetic energy inside a closed conducting box that demonstrates the resonant possibilities follows if plane waves, started in some manner in such a box, are followed in their travels. It is evident that in the general case these will be reflected continuously from the walls of the box. Certain conditions of dimensions proper compared to wavelength may exist such that standing wave patterns may be set up inside the box with constant total energy, this energy passing naturally between the electric and magnetic fields of the box. The simplest example of this may be found in a rectangular box with a plane wave bouncing between only four of the walls, as pictured in Fig. 10.02c. For the simplest case, this wave may be polarized with electric vector in the vertical or y direction and with no variations in that direction. If the path of the plane wave makes an angle θ with the normal to side 1, as shown, some general conclusions may be drawn at once from the concepts of Chapter 7 without a detailed study of the wave

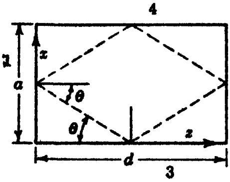


FIG. 10.02c. Paths of component uniform plane waves in a closed resonant box.

paths. It would be expected, for example, that since the vertical electric field should be zero at the conducting sides 1 and 2, the dimension d should be a half wavelength *measured at the phase velocity in the z direction.*

$$d = \frac{1}{2f\sqrt{\mu_1\epsilon_1}\cos\theta} \qquad [1]$$

where μ_1 and ϵ_1 are the constants for the dielectric filling the guide. Similarly, the dimension a should be a half wavelength measured at the phase velocity in the x direction, so that the vertical electric field may be zero at the two conducting sides 3 and 4.

$$a = \frac{1}{2f\sqrt{\mu_1\epsilon_1}\sin\theta} \qquad [2]$$

The top and bottom raise no problem since the only electric field component is vertical and so ends on top and bottom normally as required, no matter how far apart these are placed. The two conditions (1) and (2) might be combined to eliminate θ, giving

$$\omega^2\mu_1\epsilon_1 = \left(\frac{\pi}{a}\right)^2 + \left(\frac{\pi}{d}\right)^2 \qquad [3]$$

This expression shows that the natural frequency necessary to set up the assumed standing wave pattern is fixed by the dimensions a and d, and by the dielectric material filling the box. This expression will be derived in other ways in later articles, where it will be studied more completely. For the moment, it should be noted that (3) has been derived from wave solutions to Maxwell's equations and is therefore completely correct.

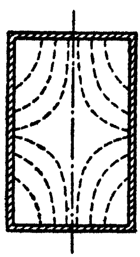


FIG. 10.02*d*. Cylindrical cavity and electric field pattern on a longitudinal section plane.

A fourth picture, quite similar to the above, suggests that a closed hollow cavity may always be considered as a wave guide with shorted ends, at least if the configuration is simple enough. Thus a conducting cylinder with closed ends as shown in Fig. 10.02*d* may be considered as a circular wave guide, and standing waves for any of the wave types of Art. 9.04 may be set up, so long as the height between the closed ends is a half wavelength (or multiple of a half wave) measured according to wavelength in the guide. It should be evident from this picture and the previous one that a particular cavity of fixed shape and size may have many different modes corresponding to all the

wave types that may exist in the corresponding wave guide, and to different integral numbers of half waves between the shorting ends. In general, the different modes will have different resonant frequencies. The wave shown in the cylinder of Fig. 10.02*d* is the standing wave corresponding to a TM_{01} wave in a circular guide. This concept of standing wave guide type waves is one which we shall develop later in studying many of the simpler cavities.

A final analogy that should not be overlooked comes from another branch of science. In the study of sound, one finds resonators for the sound waves which are quite similar to the cavity resonators for electromagnetic waves. This analogy may be appreciated from the pictures developed of the standing waves arising because of reflections of waves from the box walls. The phenomena of reflections and standing wave patterns obviously occur also for sound waves. Mathematically, the analogy is quite complete for the determination of resonant frequency since wave motion in each case is expressed by the wave equation, and certain boundary conditions must be satisfied, boundary conditions for field components at the walls in the electromagnetic case and for velocity components at the walls in the case of sound. Practically, this analogy may be of use in predicting resonant frequencies of an electromagnetic cavity resonator if the resonant frequency of a similar shaped cavity to sound waves is known.

Each of the several analogies above supplies background for understanding electromagnetic energy storage inside a hollow closed conducting box of practically any shape andfor appreciating the usefulness of this arrangement in place of the usual tuned circuit of low frequencies. It should be recognized that, except for extraneous holes or leaks that may be added in constructing the cavity practically, the region is perfectly shielded from the outside, so that there is no radiation to or interference from the outside. The behavior of the cavity for frequencies on and near resonance will be similar to that of lumped circuits with, as we shall see later, extremely high values of Q. A given cavity should have many possible modes (actually an infinite number) and for each mode the resonant frequency is determined by the mode, the cavity dimensions, and the constants of the dielectric filling the cavity. Coupling to the cavity may be either to the electric or the magnetic fields of the mode it is desired to excite.

10.03 Simple Rectangular Box Resonator

From the general discussion of the previous article, we proceed now to a more detailed study of the field, charge, and current distribution in the rectangular box resonator which was one of the simple examples of that

article. This time quantitative knowledge of the waves in rectangular wave guides will be employed and consideration given to the standing wave patterns of these waves that may be set up inside the hollow rectangular conducting box.

In this and all further resonator studies, a procedure, now quite familiar, will be followed. The waves will be studied first by assuming perfectly conducting walls for the cavities. We shall then correct for the effect of finite conductivity by calculating the losses arising from the current flow of the ideal wave in the conductors of known conductivity. This procedure was used extensively in the study of the effect of finite conductivity on propagating waves in the previous two chapters.

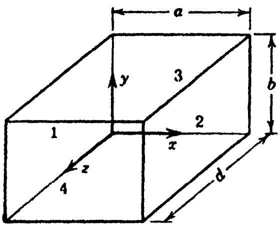


FIG. 10.03. Rectangular cavity.

One of the simplest and most useful of the wave types in a rectangular wave guide is the TE_{10} wave. This was studied in detail in Art. 9.05. If such a wave propagates in the z direction, Fig. 10.03, a resonant standing wave pattern would be expected when the dimension d is exactly a half wavelength measured at phase velocity in the guide. From Eq. 9.05(6)

$$d = \frac{\lambda_g}{2} = \frac{1}{2f\sqrt{\mu_1\epsilon_1}\sqrt{1 - (\lambda_1/2a)^2}} \tag{1}$$

where μ_1 and ϵ_1 are the constants for the dielectric material filling the box. Remembering only that

$$f\lambda_1 = \frac{1}{\sqrt{\mu_1\epsilon_1}} \tag{2}$$

(1) may be rewritten

$$\omega^2\mu_1\epsilon_1 = \left(\frac{\pi}{a}\right)^2 + \left(\frac{\pi}{d}\right)^2 \tag{3}$$

Notice that this expression is exactly that obtained by reasoning from the plane wave reflections, Eq. 10.02(3). It fixes the resonant frequency at which the desired standing wave may exist, in terms of the two dimensions a and d, and the constants of the dielectric filling the box. Notice that the dimension b does not enter into the expression, since it has no effect on the cut-off frequency of the TE_{10} wave. It may be convenient to rewrite (3) in terms of the resonant wavelength, measured according

to ve ity light in the dielectric material filling the box,

$$\lambda_1 = \frac{1}{f\sqrt{\mu_1\epsilon_1}}$$

$$= \frac{2ad}{\sqrt{a^2 + d^2}} \qquad [4]$$

Notice that for a square matic box, $a = d$, this wavelength is equal to $\sqrt{2}\,a$.

The field distributio are of great interest. Relations between field components in a sing raveling TE_{10} wave are given by Eqs. 9.05(1) to 9.05(3).

$$H_z = B \cos\frac{\pi x}{a}$$

$$H_x = jB\left(\frac{2a}{\lambda_1}\right)\sqrt{1 - (\lambda_1/2a)^2}\,\sin\frac{\pi x}{a} \qquad [5]$$

$$E_y = -j\eta_1 B(2a/\lambda_1)\sin\frac{\pi x}{a}$$

The facto $^{\omega t-\beta z)}$ is understood as multiplier of the above expressions. For a wa aveling in the negative z direction, $e^{j(\omega t+\beta z)}$, only the term in H_x has opposite sign (Art. 9.04). In each case, from Eq. 9.05(6),

$$\beta = \frac{\omega}{v_p} = \omega\sqrt{\mu_1\epsilon_1}\sqrt{1 - (\lambda/2a)^2}$$

By subs ing from (1),

$$\beta = \frac{\pi}{d} \qquad [6]$$

The of positively and negatively traveling waves may be written with es denoting the wave in the negative z direction.

ds

walls

$$H_z = (Be^{-j\beta z} + B'e^{j\beta z})\cos\frac{\pi x}{a}$$

$$H_x = (Be^{-j\beta z} - B'e^{j\beta z})j\left(\frac{2a}{\lambda_1}\right)\sqrt{1 - \left(\frac{\lambda_1}{2a}\right)^2}\,\sin\frac{\pi x}{a} \qquad [7]$$

At t $$E_y = -(Be^{-j\beta z} + B'e^{j\beta z})j\eta_1\left(\frac{2a}{\lambda_1}\right)\sin\frac{\pi x}{a}$$

In order that E_y may be maintained zero at the side $z = 0$, flected wave must be equal in magnitude and opposite in sign of E he incident wave, or by noting the last equation of (7), $B' = -E$ ith this substitution, β from (6), and the further definition,

$$E_0 = -2B\eta_1 \frac{2a}{\lambda_1}$$

the total field components (7) may be written

$$E_y = E_0 \sin \frac{\pi x}{a} \sin \frac{\pi z}{d}$$

$$H_x = -j \frac{E_0}{\eta_1} \frac{\lambda_1}{2d} \sin \frac{\pi x}{a} \cos \frac{\pi z}{d} \qquad [8]$$

$$H_z = j \frac{E_0}{\eta_1} \frac{\lambda_1}{2a} \cos \frac{\pi x}{a} \sin \frac{\pi z}{d}$$

These field distributions show first that the component of electric field E_y is zero as required on the side walls $x = 0$, $x =$ $= 0$, and $z = d$. It follows half-sine distributions between the wa nd is a maximum at the center of the box, $x = a/2$, and $z = d/2$. nce E_y is the only component of electric field, it enters the top a bottom normally, as required. The component H_x is zero at $x = 0$ $x = a$, but is a maximum at $z = 0$ and $z = d$. The reverse is true H_z, so that the magnetic field lines, if drawn out, would form cl lines surrounding the vertical displacement currents correspondi o E_y; magnetic field is always tangential to the conducting walls. re are no variations in any of the components in the vertical direction.

The charge distribution on the cavity walls is given by the ctric field ending on these:

On sides	no charge	
On bottom	$\rho_s = \epsilon_1 E_y$	[9]
On top	$\rho_s = -\epsilon_1 E_y$	

Total charge on the bottom is merely the integral of surfa arge density over the bottom; by (8) and (9),

$$q = \int_0^a \int_0^d \epsilon_1 E_0 \sin \frac{\pi x}{a} \sin \frac{\pi z}{d}\, dx\, dz$$

$$= \frac{4ad}{\pi^2} \epsilon_1 E_0$$

; fro
uenc
limen-
Notice
has no

The total charge on the top face is equal to this in magnit venient opposite sign. ccording

The current distribution is given by the components of magnetic field tangential to the conducting surfaces.

$$\text{Side 1} \quad J_y = -H_z|_{x=0} \qquad \text{Side 2} \quad J_y = H_z|_{x=a}$$

$$\text{Side 3} \quad J_y = H_x|_{z=0} \qquad \text{Side 4} \quad J_y = -H_x|_{z=d}$$

$$\text{Bottom} \quad J_x = H_z \qquad J_z = -H_x$$

$$\text{Top} \quad J_x = -H_z \qquad J_z = H_x$$

Or on sides 1 and 2
$$(J_y)_{1,2} = -j\frac{E_0}{\eta_1}\frac{\lambda_1}{2a}\sin\frac{\pi z}{d} \qquad [11]$$

On sides 3 and 4
$$(J_y)_{3,4} = -j\frac{E_0}{\eta_1}\frac{\lambda_1}{2d}\sin\frac{\pi x}{a}$$

Notice that this current flow is in the same direction on all sides at any instant. The total current flow on all sides is merely the integral of J_y over the four sides.

$$\begin{aligned} I_y &= 2\int_0^d J_{y_{1,2}}\,dz + 2\int_0^a J_{y_{3,4}}\,dx \\ &= -\frac{2jE_0\lambda_1}{\eta_1}\left[\int_0^d \frac{1}{2a}\sin\frac{\pi z}{d}\,dz + \int_0^a \frac{1}{2d}\sin\frac{\pi x}{a}\,dx\right] \\ &= -\frac{2jE_0\lambda_1}{\eta_1}\left[\frac{2d}{\pi}\cdot\frac{1}{2a} + \frac{2a}{\pi}\cdot\frac{1}{2d}\right] \\ &= -\frac{2jE_0\lambda_1}{\pi\eta_1}\left[\frac{d}{a} + \frac{a}{d}\right] \end{aligned} \qquad [12]$$

This current does not vary with the vertical dimension y along the side walls, but as it turns to flow into the distributed charge on the top and bottom, it does decrease in magnitude, falling to zero at the center of top and bottom.

The energy storage and energy loss in the box are also of fundamental interest. We may calculate exactly the energy stored in the electric fields and in the magnetic fields for the box with perfectly conducting walls. In electric fields,

$$U_E = \int_0^a\int_0^b\int_0^d \frac{\epsilon_1 E_y^2}{2}\,dx\,dy\,dz$$

At the instant of time when electric fields are a maximum, E_y may be

obtained from (8),

$$U_E = \int_0^a \int_0^b \int_0^d \frac{\epsilon_1 E_0^2}{2} \sin^2 \frac{\pi x}{a} \sin^2 \frac{\pi z}{d} \, dx \, dy \, dz$$

$$= \frac{\epsilon_1 E_0^2}{2} \cdot \frac{a}{2} b \cdot \frac{d}{2} = \frac{\epsilon_1 abd}{8} E_0^2 \qquad [13]$$

The energy stored in the magnetic fields of the box:

$$U_H = \int_0^a \int_0^b \int_0^d \frac{\mu_1}{2} (H_x^2 + H_z^2) \, dx \, dy \, dz$$

$$= \int_0^a \int_0^b \int_0^d \frac{\mu_1 E_0^2 \lambda_1^2}{2\eta_1^2 4}$$

$$\left[\frac{1}{d^2} \sin^2 \frac{\pi x}{a} \cos^2 \frac{\pi z}{d} + \frac{1}{a^2} \cos^2 \frac{\pi x}{a} \sin^2 \frac{\pi z}{d}\right] dx \, dy \, dz$$

$$= \frac{\epsilon_1 abd}{8} E_0^2 \lambda_1^2 \left[\frac{1}{4d^2} + \frac{1}{4a^2}\right] \qquad [14]$$

But by noting the value of λ_1 at resonance from (4), under this condition the maximum energy stored in the magnetic fields is exactly equal to the maximum energy stored in electric fields,

$$U_E = U_H = \frac{\epsilon_1 abd}{8} E_0^2 \qquad [15]$$

Finally we may obtain an approximation of the power loss in the side walls of the box if they are not perfect conductors. This will be done, as was stated at the beginning of this study, by assuming that fields and currents are essentially the same as for the case of perfect conductivity and by calculating the power loss due to these currents flowing in the imperfect conductors. The current distributions are given in (11). If the magnitude of surface current density, $|J|$, is known at a point, the power loss at high frequencies per unit area is $\frac{|J|^2}{2} R_s$, where R_s equals the skin effect surface resistivity. Total power loss is the integral of this quantity over all surfaces.

On sides (1) and (2),

$$P_{1,2} = 2 \int_0^b \int_0^d \frac{R_s}{2} \frac{E_0^2}{\eta_1^2} \frac{\lambda_1^2}{4a^2} \sin^2 \frac{\pi z}{d} \, dy \, dz$$

$$= \frac{R_s b d \lambda_1^2}{8\eta_1^2 a^2} E_0^2$$

On sides (3) and (4),

$$P_{3,4} = 2\int_0^a\int_0^b \frac{R_s}{2}\frac{E_0^2}{\eta_1^2}\frac{\lambda_1^2}{4d^2}\sin^2\frac{\pi x}{a}\,dx\,dy$$

$$= \frac{R_s ab\lambda_1^2}{8\eta_1^2 d^2}E_0^2$$

On top and bottom,

$$P_{T,B} = 2\int_0^a\int_0^d \frac{R_s}{2}\left[|J_x|^2 + |J_z|^2\right]dx\,dz$$

$$= \frac{R_s E_0^2\lambda_1^2}{4\eta_1^2}\int_0^a\int_0^d\left[\frac{1}{d^2}\sin^2\frac{\pi x}{a}\cos^2\frac{\pi z}{d} + \frac{1}{a^2}\cos^2\frac{\pi x}{a}\sin^2\frac{\pi z}{d}\right]dx\,dz$$

$$= \frac{R_s E_0^2\lambda_1^2}{16\eta_1^2}\left[\frac{a}{d} + \frac{d}{a}\right]$$

The total power loss,

$$P_L = P_{1,2} + P_{3,4} + P_{T,B}$$

$$= \frac{R_s\lambda_1^2}{8\eta_1^2}E_0^2\left[\frac{bd}{a^2} + \frac{ab}{d^2} + \frac{a}{2d} + \frac{d}{2a}\right] \qquad [16]$$

The simple standing wave mode inside a rectangular conducting box resonator has been investigated from the point of view of the physically real quantities: charges, currents, energies, and fields. Current and charge distribution in this mode, as for the simple mode discussed qualitatively in Art. 10.02 for a cylindrical box, is very suggestive of a lumped LC circuit extended to high frequencies. Equal and opposite charges exist on the top and bottom of the box, with charge density a maximum at the center, zero at the side walls. Current flows between these charges, vertically in the side walls, and laterally in top and bottom, decreasing to zero at the center. The electric field is vertical, passing between the opposite charges on top and bottom, and so is also a maximum at the center, zero at the edges. Magnetic fields surround these vertical displacement currents. There are no variations in any components in the vertical direction. Electric and magnetic fields are in time quadrature, and at resonance the total energy inside the resonator is a constant, interchanging between the electric and magnetic fields. Now that these facts have been established from rigorous field theory, there is a basis for drawing correct circuit analogies for resonant cavities.

Problem 10.03. Sketch lines indicating direction of *total* current in all walls of the box for the mode studied in this article.

10.04 Circuit Analogies for Simple Resonator

The fields, charges, currents, and energies studied in the last article are useful for much of our thinking about resonators. Other relations are often equally useful and are based upon conventional lumped circuit notions. In Chapter 1, the Q of a resonant circuit was useful for comparing band width of different circuits, or their excellence as energy storage devices. Also the comparison between different resonant circuits designed for a given frequency (fixed $\sqrt{LC}$) was made on the basis of the ratio $\sqrt{L/C}$. Interpreting this need in terms of the resonant cavity, it is recognized that the performance of different cavities resonant at the same frequency may be quite different. Part of this difference may be expressed in terms of a comparison between power loss and energy storage, and so it is convenient to define a Q. Also, it is not too instructive to have an expression merely for stored energy, since at resonance the maximum electric energy is exactly equal to the maximum magnetic energy, and the actual magnitude of each is dependent upon the level of excitation. However, by noting the amount of energy in terms of a representative current flowing in the resonator, an equivalent inductance may be defined, and by noting the energy in terms of a representative voltage, an equivalent capacity may be defined. These will be useful for thinking purposes, but of course it must be remembered that the mere definition of these quantities does not at once enable us to draw an equivalent circuit that will predict all properties of the resonator.

The value of Q most easily calculated is that which follows directly from the definition of Q for a lumped circuit, given in terms of energies in Art. 1.06.

$$Q = \frac{\omega(\text{Energy stored})}{\text{Average power loss}} \tag{1}$$

Both energy storage and power loss have been calculated, and we may substitute from Eqs. 10.03(15) and (16)

$$Q = \frac{\omega\epsilon_1 abdE_0^2\eta_1^2}{R_s\lambda_1^2E_0^2\left[\dfrac{bd}{a^2} + \dfrac{ab}{d^2} + \dfrac{a}{2d} + \dfrac{d}{2a}\right]}$$

Substituting from Eqs. 10.03(3) and (4) gives

$$Q = \frac{\pi\eta_1}{4R_s}\left[\frac{(a^2 + d^2)^{3/2}}{a^3 + d^3 + \dfrac{ad}{2b}(a^2 + d^2)}\right] \tag{2}$$

Note that for a cubic resonator, $a = b = d$, this reduces to the simple expression

$$Q_{\text{cube}} = \frac{2\sqrt{2}\,\pi}{12}\frac{\eta_1}{R_s} = 0.742\frac{\eta_1}{R_s} \tag{3}$$

Typical magnitudes and usefulness of this Q will be discussed later.

If it is desired to define the inductance of a cavity by calculating energy stored in terms of a representative current, it must first be decided which current should be used. The logical choice seems to be the total vertical current flowing in the side walls. Then, writing that energy stored in magnetic fields should be proportional to an inductance times the square of this current,

$$U_H = \tfrac{1}{2}L|I_y|^2$$

And substituting from Eqs. 10.03(12) and 10.03(14)

$$L = \frac{2\epsilon_1 abdE_0^2}{8}\cdot\frac{\lambda_1^2\left[\dfrac{1}{4d^2}+\dfrac{1}{4a^2}\right]\pi^2\eta_1^2}{4E_0^2\lambda_1^2\left[\dfrac{d}{a}+\dfrac{a}{d}\right]^2}$$

$$= \mu_1\frac{\pi^2}{64}\frac{abd}{a^2+d^2}\quad\text{henrys} \tag{4}$$

For a cube, $a = b = d$,

$$L_{\text{cube}} = \mu_1 a\frac{\pi^2}{128} \tag{5}$$

The choice for an equivalent capacity is possibly not so easy. For one thing, it might appear logical to define C so that the product of L and C gives resonant frequency by the conventional formula,

$$\omega^2 = \frac{1}{LC}$$

Thus, using (4) and Eq. 10.03(3),

$$C = \frac{64[a^2+d^2]\mu_1\epsilon_1}{\mu_1\pi^2 abd\pi^2\left(\dfrac{1}{a^2}+\dfrac{1}{d^2}\right)}$$

$$= \frac{64}{\pi^4}\epsilon_1\frac{ad}{b} \tag{6}$$

In addition to the above definition, at least two other alternative ones suggest themselves:

1. Capacity defined in terms of stored electrical energy and a representative voltage.
2. Capacity defined in terms of total charge on the plates and a representative voltage.

In the first of these definitions, let us use as representative voltage the maximum voltage between top and bottom, the voltage at the center of the cavity,

$$V = E_0 b$$

and setting

$$U_E = \tfrac{1}{2} C' V^2$$

from Eq. 10.03(15)

$$C' = \frac{2}{E_0^2 b^2} \epsilon_1 \frac{abd}{8} E_0^2$$

$$= \frac{\epsilon_1}{4} \frac{ad}{b} \tag{7}$$

If the same voltage is used for the second definition,

$$q = C'' V$$

q is given by Eq. 10.03(10)

$$C'' = \frac{4ad\epsilon_1 E_0}{\pi^2 E_0 b} = \frac{4}{\pi^2} \epsilon_1 \frac{ad}{b} \tag{8}$$

All the values of capacity calculated in (6). (7), and (8) are different, as would be expected, but all are proportional to $\epsilon_1 (ad/b)$, the capacity of an ideal parallel plate condenser formed by the top and bottom plates. If we denote this as C_0,

$$\begin{aligned} C &= \frac{64}{\pi^4} C_0 = 0.668 C_0 \\ C' &= \tfrac{1}{4} C_0 = 0.250 C_0 \\ C'' &= \frac{4}{\pi^2} C_0 = 0.406 C_0 \end{aligned} \tag{9}$$

If a ratio $\sqrt{L/C}$ is desired, let us use (4) and the first definition of capacity, (6)

$$\sqrt{\frac{L}{C}} = \sqrt{\frac{\mu_1 \pi^2}{64} \frac{abd}{a^2 + d^2} \times \frac{\pi^4 b}{64 \epsilon_1 ad}}$$

$$= \eta_1 \frac{\pi^3}{64} \frac{b}{\sqrt{a^2 + d^2}} \tag{10}$$

For a cube, this reduces to

$$\sqrt{\frac{L}{C}}\bigg|_{\text{cube}} = \frac{\pi^3}{64\sqrt{2}}\eta_1 = 0.342\eta_1 \qquad [11]$$

The approximate losses expressed in Eq. 10.03(16) may be expressed in two different ways. For one thing, we might ask about the impedance of the device at resonance as seen by the voltage at the center of the box. This is analogous to the information needed before a parallel resonant circuit is considered as completely described. It is known that at resonance the exciting voltage need supply no reactive energy, and the average power supplied, which may be written as proportional to a conductance and square of voltage, is equal to that lost in the box.

$$P_L = \tfrac{1}{2}GV^2 = \tfrac{1}{2}G(E_0b)^2$$

From Eq. 10.03(16),

$$G = \frac{2R_s\lambda_1^2}{E_0^2b^28\eta_1^2}E_0^2\left[\frac{bd}{a^2} + \frac{ab}{d^2} + \frac{a}{2d} + \frac{d}{2a}\right]$$

Substitute λ_1 from Eq. 10.03(4).

$$G = \frac{R_s}{\eta_1^2}\frac{\left[(a^3 + d^3) + \dfrac{ad}{2b}(a^2 + d^2)\right]}{b(a^2 + d^2)} \qquad [12]$$

It is interesting to note that this expression may be written in terms of Q and $\sqrt{L/C}$ by (10) and (2)

$$G = \frac{\pi^4}{256}\frac{1}{Q\sqrt{\dfrac{L}{C}}} = \frac{0.377}{Q\sqrt{\dfrac{L}{C}}} \qquad [13]$$

Secondly, a series resistance might be defined so that the losses at resonance are proportional to the resistance and square of the total current on the side walls.

$$P_L = \tfrac{1}{2}R|I_y|^2$$

From Eqs. 10.03(12) and 10.03(16)

$$R = \frac{R_s\pi^2}{16}\frac{b\left[(a^3 + d^3) + \dfrac{ad}{2b}(a^2 + d^2)\right]}{(a^2 + d^2)^2} \qquad [14]$$

We shall finally study typical magnitudes and the significance of the quantities defined above. From physical reasoning, it might well be

expected that the Q would be more or less proportional to the ratio of volume inside the resonator to surface area of the resonator, since the former permits energy storage, and the latter power loss. This, of course, is useful only for rough comparisons, since the field distributions should also be taken into account. However, one would predict from such a concept that Q of the simple resonator should increase with b, when b is smaller than a or d, since the ratio of volume to surface then increases. However, for very large values of b, volume and surface should increase almost proportionally. A study of (2) shows that Q is proportional to b for small values of b and independent of b for very large values of b. The order of magnitude of Q can most easily be obtained by use of (3) with air as dielectric and copper or brass as conductor. Thus for copper, $R_s = 2 \times 10^{-3}$ ohm at 1000 mc, 1.2×10^{-2} ohm at 3000 mc. The two respective values of Q are therefore 140,000 and 23,300. For brass at 1000 mc and 3000 mc, the Q is 70,000 and 11,700 respectively.

A very important question about the Q defined in terms of energy is concerned with its relation to band width. For lumped circuits, it was found that Q as defined by $\omega L/R$ gave the relation used in (1), and the band width of the resonant device was also expressed by the relation

$$\frac{\Delta f}{f_0} = \frac{1}{2Q} \qquad [15]$$

Here f_0 is the resonant frequency, and Δf is the difference between f_0 and the frequency at which impedance is $1/\sqrt{2}$ times its magnitude at resonance. In Chapter 1 it was further found that for low-loss circuits this point could be considered as that for which the reactive power supplied is equal to the loss power. Thus, if the variation in electric and magnetic energies with frequency is the same for the cavity resonator as for the circuit with lumped L and C, the value of Q defined from (1) would lead exactly to (15). This need not be true in general, but it is usually close enough over the narrow band of the resonator for the expression (15) to be useful, at least for qualitative thinking. If Q is of the order of 25,000 as suggested by previous calculation, $\Delta f/f_0$ is 0.002 per cent, or 60,000 cycles out of 3000 mc. Band width usually spoken of is twice the above value calculated from the center to one side. To increase the band width it will be necessary to decrease Q, usually by loading or adding losses.

It has been shown in previous expressions that the defined capacity of the resonator is of the same order of magnitude as the capacity of an ideal parallel plate condenser formed by the top and bottom. The ratio $\sqrt{L/C}$, to order of magnitude, is around a third of η_1, or for air dielectric,

around 100 ohms. Of course, actually it is a function of the resonator dimensions and may be varied considerably by changing these dimensions. It was also found that the shunt input impedance looking from the center of the box is of the order of $Q\sqrt{L/C}$, as in lumped circuits, and the series impedance close to $(1/Q)\sqrt{L/C}$ as in lumped series circuits. The shunt impedance is very high (order of megohms) whereas the series resistance is very low (order of a few thousandths of an ohm).

The high Q, high shunt impedance, and low series impedance each mean that the cavity resonator is an extremely efficient device for energy storage. In any practical case the losses introduced by the coupling devices may be more than those in the box itself, so that the calculated value of Q may not be obtained. However, as was stressed previously, there are no radiation losses to consider as in lumped circuits because of the practically perfect shielding afforded by the conducting walls of the cavity.

Problem 10.04(*a*). For the simple mode in the rectangular cavity studied in Art. 10.04 derive expressions for losses and Q if the loss is in the dielectric rather than the conducting boundaries.

Problem 10.04(*b*). Find the approximate new typical values for Q in the case of the copper and brass cavities mentioned in Art. 10.04 if the dielectric, instead of being air, is glass with values for ϵ_1' and ϵ_1'' of 5 and 0.05 respectively. How much is the Q changed if $\epsilon_1' = 5$ and $\epsilon_1'' = 0$?

10.05 Other Modes in Rectangular Box Resonators

It should be evident from the approach to the simple mode in a rectangular box resonator that this is only one of the many possible modes of electromagnetic oscillation that might be set up inside the box. We began by considering the rectangular box as a wave guide, requiring for oscillation the condition that a TE_{10} wave propagating in a certain direction should see a half wavelength in that direction. Obviously, any multiple of a half wave in that direction would have served as well. Thus, for a given box, the possibility exists of an infinite number of resonant frequencies corresponding to this distance equal to $p\lambda_g/2$, where p is any integer.

What now about the possibility of using other waves in the wave guide, still considering propagation along one selected direction? There should then be the above p values of resonant frequency for each of the TE_{mn} and TM_{mn} modes. Since m, n, and p may each take on integral values up to infinity, it follows that a triply infinite set of resonant frequencies corresponds to all the possible wave modes inside the box.

A certain arbitrariness exists in the selection of the direction of propagation in applying the wave guide type waves to the analysis of the

resonator. For instance, in Fig. 10.03, the waves may be regarded as TE_{10} waves, propagating in the z direction as in Art. 10.03. There are then half-wave variations in the z and x directions, no variations in the y direction. A little study should reveal that exactly the same final field pattern would be obtained if we considered this as a TE_{10} wave propagating in the x direction, setting up a standing half-wave pattern in this direction. Also, if the y direction is considered as the direction of propagation, exactly the same pattern would be found for a TM_{11} wave propagating *exactly at cut-off* in this direction. This is a possible condition for satisfying boundary conditions, since the transverse component of electric field in a TM wave is zero at cut-off. Expressed mathematically, the integer p may be zero for a TM wave when the length of the box is made equal to $p\lambda_g/2$, but it may not be zero for a TE wave.

Thus it is found that if different axes of the box are chosen for the direction of propagation, we shall again arrive at the same modes, although a given mode which looks like a certain type of guided wave for propagation along one axis may look like an entirely different type when we change to another axis.

It is of course possible to arrive at these resonant modes simply by starting directly from Maxwell's equations, asking what conditions must hold if the boundary conditions at the conductors are to be satisfied, and never referring to the study of propagating waves. The other approach has been chosen in order to make use of the background of guided wave types, and of resonance phenomena arising from interference between incident and reflected waves developed in transmission line studies.

Problem 10.05. Arrive at the resonant frequencies and field distribution patterns for the waves discussed in Art. 10.05 by starting directly from Maxwell's equations.

10.06 Simple Mode in Cylindrical Resonator

With the background developed from the detailed study of the simple mode in a rectangular box, the relations for the similar mode in a cylindrical box may quickly be set down. We should expect to find a similar mode, that is, one with the two ends charging up against each other, with axial currents flowing in the side wall, with electric fields only in the axial direction, and with no field variation in this direction. A review of the wave guide type waves for a cylinder reveals a TM_{01} wave as a likely prospect, operating in the cylinder at cut-off to insure no variation in the axial direction. The transverse component of electric field is zero at cut-off, leaving only the axial component, as desired.

Relations for the components of this wave at cut-off may be obtained from Art. 9.03.

$$E_z = E_0 J_0(k_c r)$$

$$H_\phi = \frac{jE_0}{\eta_1} J_1(k_c r) \qquad [1]$$

$$k_c = \frac{p_{01}}{a} = \frac{2.405}{a}$$

Thus the resonant wavelength is

$$\lambda_1 = \frac{2\pi}{k_c} = \frac{2\pi a}{p_{01}} = 2.61a$$

The charge density is

$$\text{On bottom} \quad \rho_s = \epsilon_1 E_z = \epsilon_1 E_0 J_0(k_c r)$$

$$\text{On top} \quad \rho_s = -\epsilon_1 E_z = -\epsilon_1 E_0 J_0(k_c r) \qquad [2]$$

Total charge, on top or bottom, is

$$q = \int_0^a \epsilon_1 E_z 2\pi r\, dr = 2\pi\epsilon_1 E_0 \int_0^a J_0(k_c r) r\, dr$$

This may be integrated by Eq. 3.22(3).

$$q = 2\pi\epsilon_1 E_0 J_1(k_c a)$$

But

$$J_1(k_c a) = J_1(2.403) = 0.5191$$

so

$$q = 2\pi(0.5191)\epsilon_1 E_0 \qquad [3]$$

The current flow on the side walls is entirely in the z direction; on the top and bottom it is radial.

$$\text{On side walls} \quad J_z = -H_\phi\big|_{r=a} = -\frac{jE_0}{\eta_1} J_1(k_c a)$$

$$\text{On bottom} \quad J_r = -H_\phi = -j\frac{E_0}{\eta_1} J_1(k_c r) \qquad [4]$$

$$\text{On top} \quad J_r = H_\phi = j\frac{E_0}{\eta_1} J_1(k_c r)$$

Total vertical current flow in the side walls is

$$I_z = 2\pi a J_z = -j\frac{2\pi a E_0}{\eta_1} J_1(k_c a) = -j\frac{2\pi(0.5191)a}{\eta_1} \qquad [5]$$

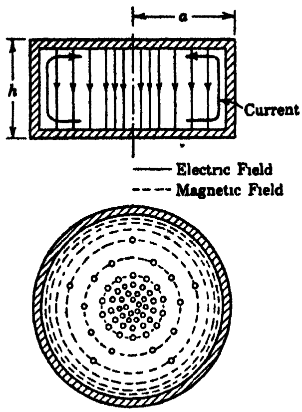


FIG. 10.06. Sections through a cylindrical cavity.

Current and field directions are sketched in Fig. 10.06.

The energy stored in the cavity at resonance may be found from the energy in electric fields at the instant these are at their maximum value.

$$U_E = h \int_0^a \frac{\epsilon_1 |E|^2}{2} 2\pi r\, dr$$

$$= \pi\epsilon_1 h E_0^2 \int_0^a J_0^2(k_c r) r\, dr$$

where h is the height of the cylinder. This may be integrated by Eq. 3.22(5).

$$U_E = \pi\epsilon_1 h E_0^2 \frac{a^2}{2} J_1^2(k_c a) \qquad [6]$$

For the resonant condition, the energy in magnetic fields at their maximum may be shown to be exactly equal to this.

If the conducting walls are of imperfect conductors, the power loss may be calculated approximately.

$$P_L = 2\pi a h \frac{R_s}{2} |J_z|^2 + 2 \int_0^a \frac{R_s}{2} |J_r|^2 2\pi r\, dr$$

The first term represents losses in the side wall, the second in top and bottom. Substituting from (4) gives

$$P_L = \pi R_s \left[a h \frac{E_0^2}{\eta_1^2} J_1^2(k_c a) + 2 \int_0^a \frac{E_0^2}{\eta_1^2} J_1^2(k_c r) r\, dr \right]$$

This integration may also be performed through Eq. 3.22(5), recalling that $J_0(k_c a) = 0$ as the condition for resonance.

$$P_L = \frac{\pi a R_s E_0^2}{\eta_1^2} J_1^2(k_c a)[h + a] \qquad [7]$$

Now we may calculate the values of Q, equivalent inductance (defined on basis of energy storage and total vertical current), equivalent capacity (defined so that resonant frequency is given by $1/(2\pi\sqrt{LC})$, the shunt conductance, and series resistance of the resonator. The method of calculation for each is of course exactly similar to that given in detail

in Art. 10.04 for the rectangular resonator. The results are:

$$Q = \frac{\omega U}{P_L} = \frac{\eta_1}{R_s} \frac{p_{01}}{2[(a/h) + 1]}$$

$$L = \frac{2U}{|I_z|^2} = \mu_1 \frac{h}{4\pi}$$

$$C = \frac{1}{\omega^2 L} = \epsilon_1 \frac{4\pi}{p_{01}^2} \frac{a^2}{h}$$

$$\sqrt{\frac{L}{C}} = \eta_1 \frac{p_{01}h}{4\pi a}$$

$$G = \frac{2P_L}{(E_0 h)^2} = \frac{R_s}{\eta_1^2} \frac{2\pi a}{h}\left[1 + \frac{a}{h}\right]$$

$$R = \frac{2P_L}{|I_z|^2} = R_s \frac{h}{2\pi a}\left[1 + \frac{a}{h}\right]$$

A study of the above relations shows the same type of behavior found for the rectangular resonator. For example, an increase in the ratio of volume to surface area increases Q as found for the rectangular box. All quantities are of the same order of magnitude, but merely multiplied by different geometric factors.

Problem 10.06(*a*). Consider a class of cavity resonators which is formed of a section of hollow pipe of arbitrary but uniform cross section closed by end plates perpendicular to the guide axis. State why, for one simple oscillation mode, the resonant wavelength is known once the cut-off wavelength of the pipe as a guide has been determined.

Problem 10.06(*b*). For the simple mode in a circularly cylindrical cavity studied in Art. 10.06 it is desired to determine the first order correction to resonant wavelength and Q when a small amount of glass is added to the cavity. Consider three cases: (1) a thin plate of glass covering one end of the cavity, (2) a thin film of glass lining the cylindrical wall portion of the cavity, (3) a thin coaxial cylinder of glass extending from top to bottom of the cavity and of diameter half that of the cavity.

10.07 Simple Mode in Spherical Resonator

A hollow conducting sphere will also have, among all its resonant modes, one which is analogous to the simple modes studied in the past articles. One pole of the sphere charges against the other, electric field passes between these equal and opposite charges, and current flows longitudinally between them, reaching a maximum value at the equator.

We shall study, somewhat later, the general wave solutions in spherical coordinates. However, the field variations required for the present

mode are relatively simple so that this general study will not be required. If no variations are assumed latitudinally, $\partial/\partial\phi = 0$, the two curl relations for the dielectric inside the hollow sphere in spherical coordinates break into two sets of equations, one relating E_ϕ, H_r, H_θ, and the other relating H_ϕ, E_r, E_θ. The mode of present interest will be obtained from the latter set. The relations from Maxwell's equations relating these are:

$$\begin{aligned} \frac{1}{r}\frac{\partial}{\partial r}(rE_\theta) - \frac{1}{r}\frac{\partial E_r}{\partial \theta} &= -j\omega\mu_1 H_\phi \\ \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(H_\phi \sin\theta) &= j\omega\epsilon_1 E_r \\ -\frac{1}{r}\frac{\partial}{\partial r}(rH_\phi) &= j\omega\epsilon_1 E_\theta \end{aligned} \qquad [1]$$

Substitution shows that the following solution satisfies equations (1).

$$\begin{aligned} H_\phi &= \frac{A\sin\theta}{k_1 r}\left[\frac{\sin k_1 r}{k_1 r} - \cos k_1 r\right] \\ E_r &= \frac{-2jA\cos\theta}{(k_1 r)^2}\left[\frac{\sin k_1 r}{k_1 r} - \cos k_1 r\right] \\ E_\theta &= \frac{j\eta_1 A\sin\theta}{(k_1 r)^2}\left[\frac{(k_1 r)^2 - 1}{k_1 r}\sin k_1 r + \cos k_1 r\right] \end{aligned} \qquad [2]$$

where $k_1 = \omega\sqrt{\mu_1\epsilon_1}$.

Assuming perfect conductivity for the spherical shell, E_θ must be zero at the radius of the shell, $r = a$. This requires, by (2)

$$\frac{(k_1 a)^2 - 1}{k_1 a}\sin k_1 a + \cos k_1 a = 0$$

or

$$\tan k_1 a = \frac{k_1 a}{1 - (k_1 a)^2} \qquad [3]$$

Roots of this transcendental equation may be found graphically or by numerical methods. The first root, other than zero, is at $k_1 a = 2.74$, or

$$\lambda_1 = \frac{2\pi}{2.74}a = 2.29a \qquad [4]$$

The representative current for the resonator may be taken where the current is a maximum, namely, across the equator.

$$\begin{aligned} I &= 2\pi a H_\phi\big|_{r=a,\,\theta=\pi/2} = 2\pi a A\left[\frac{\sin k_1 a}{(k_1 a)^2} - \frac{\cos k_1 a}{k_1 a}\right] \\ &= 2\pi a A \sin k_1 a = 0.389(2\pi a A) \end{aligned} \qquad [5]$$

The energy stored at resonance may be calculated from peak energy stored in magnetic fields.

$$U_H = \int_0^a \int_0^\pi \frac{\mu_1}{2} |H_\phi|^2 \, 2\pi r^2 \sin\theta \, d\theta \, dr$$

The value of H_ϕ is given in (1), and the result of integration may be simplified by the requirement for resonance, (3).

$$U_H = \frac{2\mu_1 \pi A^2}{3k_1^3}\left[k_1 a - \frac{1 + (k_1 a)^2}{k_1 a} \sin^2 k_1 a\right] \qquad [6]$$

The approximate dissipation in conductors of finite conductivity, having skin effect surface resistivity R_s, is

$$P_L = \int_0^\pi \frac{R_s |H_\phi|^2}{2} 2\pi a^2 \sin\theta \, d\theta$$

$$= \frac{4R_s}{3} \pi a^2 A^2 \sin^2 k_1 a \quad \text{watts} \qquad [7]$$

Since current, energy storage, and power loss are given, the circuit quantities Q, L, C, and L/C may be found from the definitions of Art. 10.04. In all these expressions, the following function of $(k_1 a)$ appears

$$F = \left[\frac{k_1 a}{\sin^2 k_1 a} - \frac{1 + (k_1 a)^2}{k_1 a}\right] = 15$$

Then

$$Q = \frac{F}{2(k_1 a)^2} \frac{\eta_1}{R_s} = \frac{\eta_1}{R_s}$$

$$L = \mu_1 \frac{aF}{3\pi (k_1 a)^3} = 0.077 \mu_1 a$$

$$C = \epsilon_1 3\pi a \frac{k_1 a}{F} = 1.7 \epsilon_1 a$$

$$\sqrt{\frac{L}{C}} = \frac{F}{3\pi (k_1 a)^2} \eta_1 = 0.21 \eta_1$$

$$R = \frac{2}{3\pi} R_s = 0.21 R_s$$

Results are again similar to those for either the cylindrical or rectangu-

lar boxes, with slightly different geometrical multiplying factors. The fields are sketched in Fig. 10.07, showing the field and current distributions described.

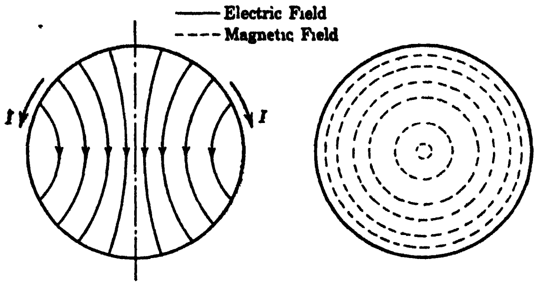


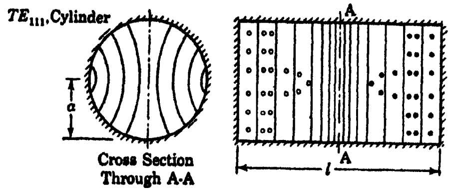
FIG. 10.07. Field patterns for simple TM_{101} mode in spherical resonator.

10.08 Miscellaneous Mode Types in Cylindrical and Spherical Resonators

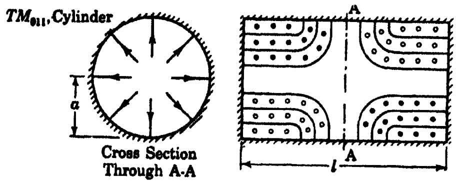
Some of the simplest modes of the many that can exist in cavity resonators have been studied in detail. Since there are an infinite number of mode types for each of the geometrical configurations of resonators, we cannot continue to give detailed information on each of these. It has already been pointed out that for a rectangular resonator, mode types will follow from each of the wave guide types studied in Chapter 9. Once the coordinate axes are chosen, any wave type may be specified by subscript, showing the number of variations in the x, y, and z directions respectively; it will be called a TM wave if it has no H_z component and a TE wave if it has no E_z. Thus a TE_{101} wave has H_z but not E_z, has one-half sine variation in the x direction, no variations in the y direction, and one in the z direction.

Similarly, wave types exist in simple cylindrical resonators corresponding to all the wave guide types of Art. 9.03. Here the commonly used order of subscripts unfortunately does not follow the cyclic order of the coordinates. Thus, the first subscript usually denotes number of ϕ variations, the second the r variations, and the third the z variations. The notation is the same as for the corresponding wave guide type, with the third subscript giving the number of axial variations. Patterns for the TM_{011}, TE_{111}, and TE_{011} wave are sketched in Table 10.08. The TM_{010} wave is that already studied in Art. 10.06. Of the additional waves noted here, the TE_{011} is perhaps the most interesting since it has no axial currents flowing in the walls, and no current flowing between the end plates and the cylindrical surface. That is, all currents are cir-

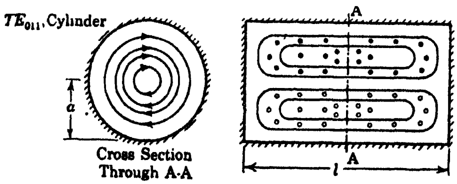
Table 10.08



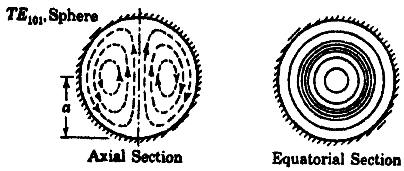
$$\lambda_1 = \frac{2l}{\sqrt{1 + \left(\frac{2l}{3.41a}\right)^2}}$$



$$\lambda_1 = \frac{2l}{\sqrt{1 + \left(\frac{2l}{2.61a}\right)^2}}$$



$$\lambda_1 = \frac{2l}{\sqrt{1 + \left(\frac{2l}{1.64a}\right)^2}}$$



$$\lambda_1 = 1.40a$$

cumferential. Thus if a resonator for such a wave is tuned by moving the end plate, one does not need to worry here about a good contact between end plates and the cylinder. For most wave types this is an important point since a large current usually flows between the cylinder and its ends.

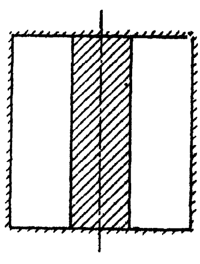


Fig. 10.08. Cylindrical resonator with dielectric core.

Of the many waves of a spherical resonator, probably the one of major interest other than that of Art. 10.07 has field distribution as sketched in Table 10.08, labeled TE_{101} for a sphere. There are only components E_θ, H_r, and H_θ for this resonator, and again no variations with azimuthal angle. Current flows latitudinally here instead of longitudinally as in the other wave type studied. This wave type may be denoted as a TE_{101}, and the one

studied previously as a TM_{101}, the subscripts denoting r, ϕ, and θ variations respectively, and the classification TM or TE wave being fixed by the restriction of magnetic or electric components to the directions transverse to the radial, namely, the ϕ and θ directions.

Problem 10.08. Find the resonant wavelength for a cavity consisting of a section of circular cylinder terminated by two end plates perpendicular to the axis and containing a coaxial dielectric core (Fig. 10.08). Consider only the simple mode in which the top and bottom are equally and oppositely charged and the current flow is symmetric about the axis.

SMALL-GAP CAVITIES AND CAVITY COUPLING

10.09 Foreshortened Coaxial Line Resonators

The resonators of simple geometrical shape — rectangular, cylindrical, spherical — have been studied so that we might become familiar with the phenomenon of resonance in closed conducting cavities, obtain some feeling for the relations between these and more familiar resonant circuits, and have some practice in manipulating some of the simpler wave solutions. Unfortunately, these simple resonators are often not so useful as some which are more difficult to analyze, at least if highly exact results are desired. A more useful class of resonators for many applications is that in which the region of maximum electric field appears across a short gap, so that the resonator may be used across two closely spaced electrodes of typical high-frequency electron tubes. For example, the resonator may be excited by a velocity modulated or a current modulated beam of electrons passing across this gap. This would be true only for the rectangular box or cylindrical box if the height were made as small as desired for the gap, but we have found that such proportions would result in a resonator of extremely low Q.

The resonators with short gaps may have many configurations and proportions. One of the simplest of these is as shown in Fig. 10.09*a* where the resonance region is located between two coaxial conductors, shorted at one end and terminated by a short gap between two conducting plates at the other end. This is the gap across which a voltage may be set up to excite a beam of electrons, or conversely, across which the modulated beam of electrons may pass to induce current flow into the resonator. Provided that the region B, in the vicinity of the gap, is short compared to a wavelength, this region will act mainly as a lumped capacity loading on a transmission line represented by the coaxial region A. Thus the equivalent circuit is approximately as in Fig. 10.09*b*. For resonance, the impedance looking into the short-circuited transmission line should be an inductive reactance equal in magnitude to

the capacitive reactance of the lumped capacity C_0. (To be more general, it is worth noting that the line might be cut any place we please, and if it is to be a self-contained system at resonance, the impedances looking in opposite directions should be equal and opposite reactances, neglecting losses.)

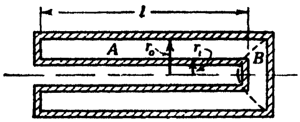


FIG. 10.09*a*. Foreshortened coaxial line resonator.

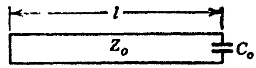


FIG. 10.09*b*. Approximate equivalent circuit for Fig. 10.09*a*.

The input impedance of a short-circuited transmission line of length l, characteristic impedance Z_0, is

$$Z_i = jZ_0 \tan \beta l \tag{1}$$

This impedance should be equal and opposite to that of the condenser C_0,

$$Z_c = \frac{-j}{\omega C_0} \tag{2}$$

$$Z_0 \tan \beta l = \frac{1}{\omega C_0}$$

or

$$\beta l = \tan^{-1} \left(\frac{1}{Z_0 \omega C_0}\right) \tag{3}$$

For a coaxial line (Table 9.01)

$$Z_0 = \frac{60}{\sqrt{\epsilon_1'}} \ln \left(\frac{r_0}{r_i}\right) \text{ ohms} \quad \beta = \frac{2\pi}{\lambda_1}$$

For a given characteristic impedance Z_0 and end loading capacity C_0 the length of the coaxial line required for resonance is fixed. If C_0 is so small that $1/\omega C_0$ is much greater than Z_0, $\beta l \cong \pi/2$, and the line is practically a quarter wave in length. For larger values of C_0, the line is foreshortened from the quarter-wave value because of this loading.

The above analysis is only approximate, but is very simple and fortunately quite useful for many practical cases. The criteria for usefulness should be:

1. The region B small compared with wavelength.
2. The region A long enough so that the doubt in the point to which l should be measured is unimportant.

The method is particularly useful when the region B is not uniform, but varies in spacing, and contains glass or other dielectric discontinuities, so long as a reasonably good value for C_0 of this space may be estimated.

Problem 10.09(*a*). Prove the statement made above that for resonance in a closed system, impedance at any point should represent opposite reactances looking in opposite directions. Show that the same result is obtained for resonance if the line is cut at a general point distance x from the condenser C_0, and this criterion applied.

Problem 10.09(*b*). Calculate the Q of the resonator of Fig. 10.09*a*, neglecting all losses in the region B.

Problem 10.09(*c*). Repeat Prob. 10.09*b* if losses in region B may be considered as a lumped resistance R_0 in parallel with C_0. Design and calculate the magnitude of Q for a resonator with $r_i = 1$ cm, $r_0 = 3$ cm, $\lambda = 30$ cm, $C = 2\ \mu\mu fd$, $R_0 = 10{,}000$ ohms.

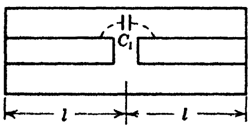


FIG. 10.09*c*.

Problem 10.09(*d*). By extension of the concepts of this article, show that the expression for resonant frequency for the resonator of Fig. 10.09*c*, having total gap capacity C_1, is

$$\beta l = \tan^{-1}\left(\frac{1}{2Z_0\omega C_1}\right)$$

10.10 Foreshortened Radial Lines

In Art. 9.08 the transmission line type of wave propagating radially between two parallel circular plates was studied. For this wave, equal and opposite currents flow radially in the top and bottom plates at a given radius, and equal and opposite charge distributions exist on these plates. Electric field is only in the axial direction, magnetic field in the circumferential direction. These field distributions are reminiscent of those for the simple cylindrical resonator, Art. 10.06, and a little study of the radial transmission line equations, Art. 9.08, reveals that such a resonator is truly the radial analogy of a shorted quarter-wave transmission line. That is, looking radially outward from the center, an infinite impedance is seen at the center. (For the simple mode of Art. 10.06 it was found that current is zero at the center, voltage is a maximum.) Equation 9.08(14) gives the input wave impedance of a short-circuited radial line

$$Z_i = jZ_{0i}\frac{\sin(\theta_i - \theta_L)}{\cos(\psi_i - \theta_L)} \tag{1}$$

This impedance is infinite if $\psi_i - \theta_L$ is $\pm 90°$, or an odd multiple of $90°$. Since $r_i = 0$ at the center, $\psi_i = 0$ and a value of $\theta_L = 90°$ fulfils the requirement for infinite impedance. By the curve, Fig. 9.08*b*, this

occurs for a value of $kr_2 = 2.4$, of course the same result for resonance obtained in Art. 10.08.

It follows that a resonator with a small gap at the center, similar to that of Art. 10.09 but of different proportions so that radii are greater than height, is best looked upon as a radial line foreshortened by capacity loading, Fig. 10.10*a*, rather than a coaxial line as in Art. 10.09. The calculation for the loaded coaxial line was fortunately quite simple. The corresponding calculation for the radial line may be a bit more difficult, but that is no excuse for neglecting the foreshortening, even though the student may use the old excuse, "I didn't know it was loaded." Actually, the amount of foreshortening by a given capacity may be calculated easily enough through the use of the radial transmission line equations. As for the coaxial line, the impedance of the shorted radial line, looking in at r_1, should be an inductive reactance equal in magnitude to the capacitive reactance of the lumped capacitance. Equation (1) for wave impedance of a shorted radial line may be written for this impedance at r_1, if the line is shorted at r_2,

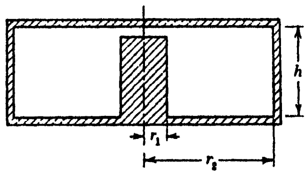


FIG. 10.10*a*. Foreshortened radial line resonator.

$$Z_1 = jZ_{01} \frac{\sin(\theta_1 - \theta_2)}{\cos(\psi_1 - \theta_2)}$$

This quantity represents the ratio of E to H at r_1. The ratio of voltage to current, or total impedance, looking outward from r_1 is

$$Z_{T_1} = -\frac{h}{2\pi r_1} \frac{E}{H}\bigg|_1 = -j\frac{hZ_{01}}{2\pi r_1} \frac{\sin(\theta_1 - \theta_2)}{\cos(\psi_1 - \theta_2)} \quad [2]$$

Then, for resonance,

$$\frac{1}{\omega C} = -\frac{h}{2\pi r_1} Z_{01} \frac{\sin(\theta_1 - \theta_2)}{\cos(\psi_1 - \theta_2)} \quad [3]$$

Most often, the value of r_1, h, and C are known, and it is desired to find the outer radius r_2 for resonance at a given wavelength.

Let

$$\frac{2\pi r_1}{\omega C Z_{01} h} = q \quad [4]$$

If this quantity is substituted in (3), the resulting equation may be solved for θ_2

$$\theta_2 = \tan^{-1}\left(\frac{\sin\theta_1 + q\cos\psi_1}{\cos\theta_1 - q\sin\psi_1}\right) \qquad [5]$$

Once θ_2 is found, $2\pi r_2/\lambda$ is read from Fig. 9.08*b*.

Problem 10.10(*a*). Plot a curve of r_2 versus λ for a radial cavity of height 1 cm, loaded by a post 1 cm in diameter, which represents a capacitance of 1 $\mu\mu fd$.

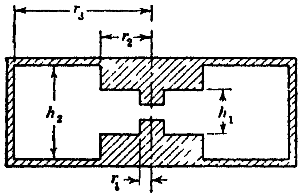


FIG. 10.10*b*. Axially symmetric resonator.

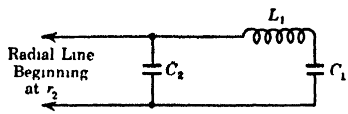


FIG. 10.10*c*. Approximate equivalent circuit for Fig. 10.10*b*.

Problem 10.10(*b*). A radial cavity is loaded at the center by a section as shown in Fig. 10.10*b*. If r_2 is relatively small compared with wavelength, it is possible to represent approximately the region inside r_2 by a lumped circuit equivalent, Fig. 10.10*c*. Here C_1 is the center post capacitance, L_1 is an inductance calculated from D-C formulas for the coaxial region of height h_1 between radii r_1 and r_2, and C_2 is approximately the capacitance calculated on the basis of parallel disks spaced h_1, and of radii r_1 and r_2. If $C_1 = 1\ \mu\mu fd$, $h_1 = 0.5$ cm, $h_2 = 1.0$ cm, $r_1 = 0.50$ cm, $r_2 = 1.0$ cm, find the approximate value of r_3 for resonance at $\lambda = 15$ cm.

10.11 Transition between Coaxial and Radial Resonators

A study of the figures for the two resonators, foreshortened coaxial in Art. 10.09, and foreshortened radial in Art. 10.10, shows that these are really not different in physical configuration except in regard to proportions. However, it was implied that in order for a cavity to be thought of usefully as a foreshortened coaxial line for the simple mode, the length l, Fig. 10.09*a*, should be relatively long compared with $(r_2 - r_1)$; in order for the radial line point of view to be useful, the height h of Fig. 10.10*a* (corresponding to l above) should be relatively short compared with $(r_2 - r_1)$. The significance, electromagnetically, is that in the former the electric field lines in the region A will tend to become radial lines, passing between inner and outer conducting cylinders, whereas in the latter they will tend to become axial lines, passing between the top and bottom plates. This is illustrated in Figs. 10.11*a* and *c*. The magnetic field lines in either case are circumferential, but in the former case their strength varies with z, whereas in the latter it does not.

The above remarks suggest that we have not studied two different wave types but rather two different approximate analyses for a given wave type, the usefulness of each depending upon the resonator proportions. As the proportions are gradually changed from Fig. 10.11*a* toward Fig. 10.11*c*, the field distribution will also gradually change from one tendency toward the other. An intermediate case is sketched in Fig. 10.11*b*. For such an intermediate case, it may be necessary to consider a more rigorous attack than either of the approximate analyses provides, and one such attack will be reviewed in the following article.

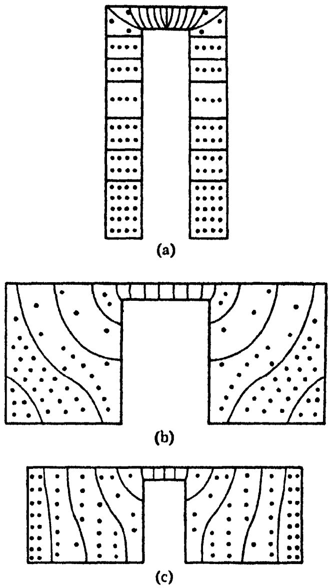


FIG. 10.11. Transition between foreshortened coaxial line and foreshortened radial line.

Before leaving the general comparison of the two limiting cases, it is especially interesting to note the limit of each as the capacity loading is increased to a very large value. For both types of resonators, the design equations show that the resonator region of the coaxial or radial line external to the lumped capacity decreases in size as the capacity loading increases for a fixed frequency. When loading is so great that the overall size of the resonator is relatively small compared with wavelength, it is permissible to consider the outside portion of line as a lumped inductance, whose value may be calculated from the formula for inductance of a coaxial system (Table 9.01).

$$L = \frac{\mu_1 l}{2\pi} \ln\left(\frac{r_2}{r_1}\right) \quad \text{henrys} \tag{1}$$

Then resonant frequency is given by the familiar equation

$$f = \frac{1}{2\pi\sqrt{LC}} \tag{2}$$

These are quite obviously the equations we would have written for this case of dimensions small compared with wavelength if we had not been exposed to resonant cavity theory. However, it is interesting as a connection between the resonant cavities and resonant lumped circuits. It is also significant in pointing out that a closed, conducting resonant region may be made indefinitely small compared with wavelength if the

fields are concentrated in this fashion. Recall that for the *simple* resonators it was found necessary to have at least some dimension comparable to wavelength. This cautions against using dimensions based upon those simple resonators in seeking estimates of dimensions for cavities in which there are highly concentrated regions of electric or magnetic field. Of course, if a very small resonator were made in this fashion it would have a low ratio of L/C and a low Q because of its poor volume to surface ratio. It would consequently be a relatively poor resonator for many purposes.

Problem 10.11(*a*). Show that Eqs. 10.11 (1) and (2) are the limiting equations, as C is made very large, for the results of Art. 10.09.

(*b*). Repeat for Art. 10.10.

10.12 The Nature of an Exact Solution

For resonators of the small-gap type, it is possible to obtain exact solutions for certain shapes such as the two examples of Figs. 10.12*a* and *b*. Such solutions may be desirable when proportions are such that approximate results from either Art. 10.09 or Art. 10.10 do not represent particularly good approximations. Although space does not permit the development of the exact method here, Hahn[2] has given one form of

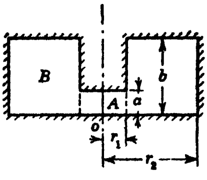


Fig. 10.12*a*.

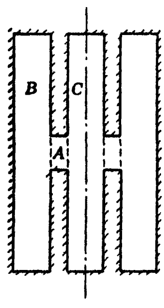


Fig. 10.12*b*.

this in the literature. The basis for the method is, however, similar to that discussed previously in connection with other problems, and especially to the discussion of discontinuities in lines and guides, Art. 9.15. Separate solutions to the wave equation are thus added to form a series which matches the somewhat complex boundary that could not be satisfied by one simple function alone.

Consider specifically the resonator of Fig. 10.12*a*. If the mode is

[2] W. C. Hahn, "A New Method for the Calculation of Cavity Resonators," *Journ. Appl. Phys.*, **12**, 62–68 (January, 1941).

desired with cylindrical symmetry and the high voltage region appearing across the gap, study shows that the field components E_r, E_z, and H_ϕ only are required. Thus in region A all possible cylindrically symmetric wave solutions with these components are written, not forgetting the important component corresponding to the radial transmission line mode (Art. 10.10). Of course, the second type (N_0) Bessel functions are absent in the A region solutions due to the axis being contained in this region, and the requirement that $E_r = 0$, at $z = 0$ and $z = a$, is also placed on these solutions. Similarly, all possible cylindrically symmetric wave types in the B region with E_r, E_z, and H_ϕ, again including the radial transmission line mode, are written with E_r made zero at $z = 0$ and $z = b$, and $E_z = 0$ at $r = r_2$ for all waves. The final matching conditions are the requirements that *total* E_r and H_ϕ from the A wave solutions be equal to that from the B wave solutions at all points along the surface $r = r_1$ from $z = 0$ to $z = a$. By means of Fourier series, and certain summations tabulated by Hahn, the coefficients of the individual waves may be evaluated and resonance conditions found.

As in Art. 9.15, it is possible to express the effect of the higher order waves (those other than the radial transmission line mode) as a lumped admittance placed at $r = r_1$. Thus the analysis can be made from an equivalent radial transmission line circuit, with a radial line corresponding to region B joined directly to a radial line corresponding to region A, with the lumped admittance shunted across the junction at $r = r_1$. Although exact calculation of this admittance requires solution of the problem outlined above, its order of magnitude may be estimated from the curve of Fig. 9.15b:

$$C_d \sim 2\pi r_1 C'_d \tag{1}$$

C_d is a lumped capacitance to place at the junction between radial lines A and B, and C'_d is obtained as a function of a/b from Fig. 9.15b.

10.13 Conical Line Resonators

The resonators of Arts. 10.09 and 10.10 may be made with relatively high Q's, yet with a small gap which can be excited conveniently by an electron stream. There are many other physical configurations of resonators which will accomplish the same result. One of the most interesting of these is the conical analogy to the coaxial and radial line cavities already studied. This is shown in Fig. 10.13 where two coaxial, conical conducting surfaces with apices adjacent are terminated at radius a by a spherical conducting surface. The region inside the conductors is filled with a dielectric of constants μ_1 and ϵ_1.

The basic equations for the principal wave guided by two coaxial cones

were presented in Art. 9.09. It was shown that this wave is exactly analogous to a transmission line wave along a uniform two-dimensional system. Current flows radially along the two cones; voltage difference between the two cones at any radius may be talked about conveniently since there is only a θ component of electric field, and the wave propagates radially at the velocity of light for the dielectric material surrounding the cones. The characteristic impedance is a constant, independent of radius (unlike that for the radial lines) and is given by Eq. 9.09(9).

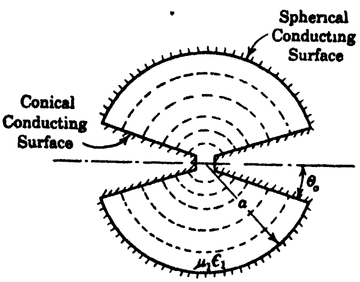


FIG. 10.13. Conical line resonator.

$$Z_0 = \frac{\eta_1}{\pi} \ln \cot \frac{\theta_0}{2} \qquad [1]$$

Ordinary transmission line equations may now be used with this value of Z_0 and the known velocity of propagation, at least if losses are neglected. Then, if the lumped capacity at the point where the two cones approach one another is negligible, resonance will occur when the spherical surface is removed by a quarter wavelength from the center so that an infinite impedance is seen looking outward from the center.

$$a = \frac{\lambda_1}{4} = \frac{1}{4f\sqrt{\mu_1\epsilon_1}} \qquad [2]$$

The field equations show that this is the exact resonance condition for the ideal case of the two cones coming to point apices, with these apices removed from one another by only an infinitesimal distance. (They must not be touching since the maximum voltage between cones appears at this point.) These field distributions for E and H may be obtained by superposing a radially outward traveling wave and a radially inward traveling wave of equal magnitudes and phases [Eqs. 9.09(5) and 9.09(6)] so that E_θ is maintained zero at $r = a$. Then,

$$\begin{aligned} E_\theta &= \frac{C}{\sin\theta}\,\frac{\cos k_1 r}{r} \\ H_\phi &= \frac{C}{j\eta_1 \sin\theta}\,\frac{\sin k_1 r}{r} \end{aligned} \qquad [3]$$

where $k_1 = \omega\sqrt{\mu_1\epsilon_1}$.

The integrations for energy, losses, and the circuit definition for resona-

tors give the following results.

$$f_\theta = \ln \cot \frac{\theta_0}{2}$$

$$I = \frac{2\pi C}{j\eta_1} \quad \text{(maximum current at the spherical boundary)} \tag{4}$$

$$U = \epsilon C^2 \pi a f_\theta \quad \text{(peak stored energy)} \tag{5}$$

$$P_L = \frac{2\pi R_s C^2}{\eta_1^2}\left[f_\theta + \frac{0.825}{\sin\theta_0}\right] \quad \text{(average power loss)} \tag{6}$$

$$Q = \frac{\eta_1 \pi}{4R_s} \frac{f_\theta}{f_\theta + \dfrac{0.825}{\sin\theta_0}} \tag{7}$$

$$L = \frac{2U}{|I|^2} = \frac{\mu_1 a}{4\pi} f_\theta \tag{8}$$

$$V = 2Cf_\theta \quad \text{(voltage between apices)} \tag{9}$$

$$G = \frac{2P_L}{V^2} = \frac{\pi R_s}{\eta_1^2} \frac{f_\theta + \dfrac{0.825}{\sin\theta}}{f_\theta^2} \tag{10}$$

$$R = \frac{2P_L}{|I|^2} = \frac{R_s}{\pi}\left[f_\theta + \frac{0.825}{\sin\theta}\right] \tag{11}$$

If the capacity between the two apices is not negligible, because of a flattening of the apices, introduction of glass, etc., the resonator will, of course, have to be foreshortened to compensate for this, just as for the coaxial and radial cavities with lumped capacity loading. If the magnitude of the lumped capacity at this point is estimated as C_0, the approximate resonance condition is written in exactly the same form as for the coaxial lines, Eq. 10.09(3),

$$\beta a = \tan^{-1}\left(\frac{1}{Z_0 \omega C_0}\right) \tag{12}$$

where Z_0 is now given by (1) and $\beta = 2\pi/\lambda_1 = \omega\sqrt{\mu_1\epsilon_1}$.

Problem 10.13(*a*). Find the values for θ_0 that will lead to maximum Q and minimum G, Eqs. 10.13(7) and 10.13(10), for the conical cavity.

Problem 10.13(*b*). Design a conical resonator for $\lambda = 15$ cm with the angle for maximum Q. Calculate this Q and the value of G if the conductor is copper.

10.14 Coupling to Cavities

The types of electromagnetic waves that may exist inside closed conducting cavities have been discussed without specifically analyzing ways of exciting these oscillations. Obviously they cannot be excited if the resonator is completely enclosed by conductors. Some means of coupling electromagnetic energy into and out of the resonator must be introduced from the outside. Some of these coupling methods have been implied in past articles. All are similar to those discussed in Art. 9.11 for exciting waves in wave guides. The most straightforward methods are:

1. Introduction of a conducting probe or antenna in the direction of the electric field lines, driven by an external transmission line.
2. Introduction of a conducting loop with plane normal to the magnetic field lines.
3. Introduction of a pulsating electron beam passing through a small gap in the resonator, in the direction of electric field lines.

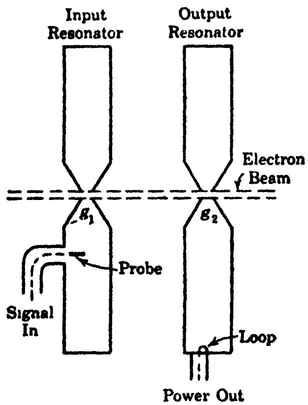


FIG. 10.14*a*. Couplings to the cavities of a velocity modulation tube amplifier.

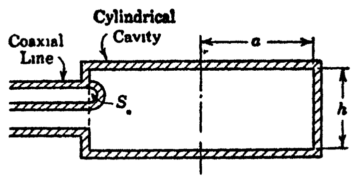


FIG. 10.14*b*. Magnetic coupling to a cylindrical cavity.

Any of the above methods may also be used for coupling power out of the resonator, as well as for introducing power to excite the oscillations in the resonator, and it is not uncommon to find all the methods employed in a single system. For example, in a velocity modulation device of the Klystron type, as pictured in Fig. 10.14*a*, the input cavity may be excited by a probe, the oscillations in this cavity producing a voltage across gap g_1 and causing a velocity modulation of the electron beam. The velocity modulation is converted to conduction current modulation by a drifting action so that the electron beam may then excite electro-

magnetic oscillations in the second resonator by passing through the gap g_2.[3] Power may be coupled out of this resonator by a coupling loop and a coaxial transmission line.

The qualitative performance of the above types of coupling is not hard to understand, but quantitative analysis is more difficult, and in certain respects has not yet been worked out rigorously. We shall consequently discuss the problem only from an approximate point of view. Let us consider specifically the case of a cavity excited by a loop, and say, for example, that this loop is introduced near the side wall of a cylindrical cavity, Fig. 10.14*b*. If a current is made to flow in this loop, all wave types will be excited which have a magnetic field threading the loop. Certainly the simple wave type of Art. 10.06 is one of these since it has only a circumferential magnetic field which is a maximum at the cylindrical wall of the cavity, and which does not vary in the axial direction. However, there are an infinite number of other wave types which have a circumferential magnetic field to thread the loop. If one of these is near resonance, and the others far from resonance, the one nearest resonance will, of course, be excited the most. Those far from resonance will be highly reactive and so will be excited little.

Suppose now the cylinder is designed with dimensions such that the simple wave type of Art. 10.06 is near resonance for the operating frequency. This wave type will fit the boundary conditions imposed by the perfectly conducting box alone. When the conducting loop is introduced, there must be some revision in the field distribution because of this loop since the electric field must also be made zero along the conductor of this loop (assuming the loop to be perfectly conducting). We may think of this as accomplished by superposing higher order waves on the main wave until the required boundary condition is met. These higher order waves will be, in general, far from their individual resonance points, and so will present only reactances to the driving source. Now, from a circuit point of view, part of the problem is to find the impedance seen by a voltage applied to the loop, or the current which flows for that voltage. For a given applied voltage, there must be an equal and opposite induced voltage, if the resistance component of the loop is negligible. This induced voltage is generated by changing magnetic flux enclosed by the loop from all the waves, simple and higher order. All the contributions from the latter, as we have seen, will be reactive and far from resonance so that the contribution to induced voltage from all these may be thought of as the inductive reactance of the loop, of course taking account of the nearness of the surrounding conductor.

[3] Brainerd, Koehler, Reich, and Woodruff, "Ultra-High-Frequency Techniques," Van Nostrand, 1942, Chapter X.

The induced voltage from the wave near resonance may be calculated by integrating magnetic flux density for this wave over the area of the loop, field distributions being already known. If the loop is small compared with dimensions of the resonator, fields will not vary greatly over the area of the loop, so that magnetic field may be assumed constant over that area. Induced voltage for this wave with field strength H at the position of the loop is then

$$V_1' = -j\omega\mu_1 HS \tag{1}$$

where S is the area of the loop. Total induced voltage is the sum of this and the reactance term representing higher order waves, and applied voltage is equal and opposite to induced voltage.

$$V_0 = j\omega\mu_1 HS + j\omega LI_L \tag{2}$$

I_L is the loop current and L is the self inductance of the loop in the presence of the shield, representing the reactive effects of all higher order waves. At resonance, no reactive energy has to be supplied to the simple wave, but only the real energy for the losses in the cavity for this wave. Thus the first term in (2) should represent a voltage in phase with current of an amount necessary to supply those losses. An expression for power loss is given in

$$I_L(j\omega\mu_1 HS) = \text{Power loss} = P_L$$

From Art. 10.06 we may find H and P_L for the TM_{010} wave in the cylindrical resonator.

$$H = H_\phi\Big|_{r=a} = \frac{jE_0}{\eta_1} J_1(k_c a)$$

$$P_L = \frac{\pi a R_s E_0^2}{\eta_1^2} J_1^2(k_c a)(h + a) = -\pi a(h + a)H^2 R_s$$

so

$$j\omega\mu_1 HSI_L = -\pi a(h + a)H^2 R_s$$

or

$$H = \frac{-j\omega\mu_1 S}{\pi a(h + a)R_s} I_L \tag{3}$$

From (3) H may be obtained in terms of I_L, so that the level of excitation of the wave in the cavity is known once the current in the loop is known. Also, by substituting in (2), the impedance looking into the loop is given. The first term represents the impedance coupled into the

loop by the resonator, and the second term represents the self inductance of the loop.

$$Z_L = \frac{V_0}{I_L} = \frac{(\omega\mu S)^2}{\pi a(h + a)R_s} + j\omega L \qquad [4]$$

We have already discussed a little of the problem of exciting the wave off resonance (Art. 10.04). The above expressions of course apply only for resonance in the main wave, and for other frequencies the manner in which electric and magnetic energies vary with frequency would have to be considered since the difference between these two terms must be supplied as a reactive energy term in addition to that lumped as the loop inductance.

The above approach was intended to clarify the mechanism of coupling between the loop and the wave in the resonator, and to supply an approximate means of analysis. It can be extended in an obvious manner for coupling by loops to other wave types and other resonator shapes; it can be extended in a somewhat less obvious manner to coupling by probes and antennas. In the simple form described here this coupling calculation has been used successfully for practical cavity design from the time the problem of cavity coupling first came up.

Problem 10.14(a). A cylindrical cavity is excited at the resonant frequency for its simple (TM_{010}) mode by a loop in one side of the resonator, and power is taken out of a loop of equal size on the other side of the resonator. Set up an approximation to the coupling problem between input and output. Calculate the current to a 70-ohm resistance load across the output when a voltage of 100 volts is impressed across the input loop. The loops are 1 cm^2 area. The resonator is of copper, 10 cm in diameter and 3 cm high.

Problem 10.14(b). Discuss the problem of coupling to the TM_{010} mode in a cylindrical resonator by a small probe or antenna, covering the points discussed in this article for a loop coupling.

Problem 10.14(c). Write the expression similar to Eq. 10.14(4) for a rectangular and spherical resonator (Arts. 10.03 and 10.07).

11

RADIATION

11.01 The Problems of Radiation Engineering

Radiation of electromagnetic energy, to an engineer, is important in at least two cases. (1) It may be a desired end result if energy is to be transferred from a high-frequency transmitter to electromagnetic waves in space by means of some antenna system. (2) It may be a leakage phenomenon, adding undesired losses to an imperfectly shielded circuit or transmission line or to a cavity resonator with holes.

In order to perform an intelligent job of engineering in either of the above radiation problems, it is first desirable to have a good physical picture of radiation. In this picture radiation is not a mysterious and unknown link between transmitter and receiver, but a phenomenon following naturally from the excellent pictures of wave propagation, reflection, and excitation built up from familiarity with transmission lines and wave guides. It is desirable that this physical picture be concrete enough to give qualitative answers to specific questions that may arise in either of the above roles of radiation. It is, of course, also necessary to have methods available for obtaining quantitative design information about the amount of radiation and the effects on the radiating system. If radiation is the desired product, several or all of the following problems may arise in design of the radiating system:

1. The field strength at a known distance and in a known direction from the radiator excited by a given voltage may be desired. Often the relative field strength versus direction, i.e., the directivity pattern, is a sufficient answer for this problem.
2. The total power radiated from the antenna structure when excited by a known voltage or current may be desired. (The answer may often be expressed in terms of a radiation resistance.)
3. The input impedance of the radiator to the exciting voltage or current may be desired.
4. The resonant frequency and band width of the radiator may be required. Band width questions are often answered if impedance versus frequency is known; however, it may be necessary sometimes to know the change in the radiation pattern with frequency.
5. The power dissipated in ohmic losses in the radiator, as compared to the power radiated, may be desired. The result may be expressed as a radiation efficiency.

6. The value of maximum gradient along the antenna may be required if corona difficulties are important.

If radiation is the leakage product, the problems are not essentially different, although a knowledge of power lost by radiation is usually sufficient. To assure ourselves that it is only in magnitude of importance that this differs from radiation as a desired product, it may be recalled that it is in the role of a leakage phenomenon that radiation was met previously. In Chapter 5, for the rigorous study of circuits, an energy loss term appeared which was not accounted for by ohmic dissipation, this term being the radiated energy. The term becomes more important as the circuit is made large compared with wavelength, suggesting the obvious conclusion that a well-designed antenna system is simply a circuit made purposely large compared with wavelength to increase the importance of radiation. Also, in the study of transmission lines, it was pointed out that the waves excited in space by the end effects of a transmission line, required for matching to these end effects, may take energy from the guided wave of the line. This too is radiation, and to obtain it as a major effect it is necessary only to accentuate these end effects or to match more closely to the waves in space. This latter point of view is excellent, and one that will be developed further.

Of the six quantitative problems listed above, it is fortunate that methods for calculating satisfactory answers to two of the most important, directivity pattern and total power radiated, have been available for many years. The calculations are usually not of prohibitive difficulty, although they may require graphical integration for many antenna configurations. We shall see that the possibility of obtaining good engineering answers to these quantities comes about for the very fortunate reason that those two quantities are relatively insensitive to small changes in current distributions over the radiating system. Consequently, with a little experience, some very good approximations to current distributions may be assumed throughout the radiator; from the integral forms of Maxwell's equations (Chapter 4) the fields at any point may be obtained if currents throughout the system are given, and through Poynting's theorem energy flow may be calculated once fields are known. (We shall also see that by a direct extension of the above point of view, fields at any point, and consequently energy radiated, may be calculated if field distributions are given over some surface surrounding the radiator. For certain types of radiating systems we can do a better job of assuming fields than currents, so the assumed fields will make a good starting point for calculating answers to the first and second problems in such cases.)

The fifth problem, power dissipation, has also been in fair shape for some time, simply because a satisfactory first approximation for many

antenna systems is: power dissipated negligible compared with power radiated. Then a quite good second approximation may be found by integrating losses due to an assumed current distribution over the known conductors of the radiating system. This will, of course, be the same current distribution assumed for calculation of power radiated if such a method is used.

The information of input impedance is important since the radiator must be matched to some transmitter or transmission line, and if one is interested in wide-band radiation, as in television or frequency modulation, it is desirable to know how good this match may be over the entire band width. Until recently such information was gathered mostly by experience, augmented by certain design curves, such as those of Siegel and Labus,[1] which give satisfactory answers for a certain class of antennas, but which were based upon a development involving many assumptions difficult to evaluate. Recently Stratton and Chu[2] have obtained, by means of exact solution of the differential equations subject to boundary conditions, results for spheroidal antennas which may be considered as good approximations to many antennas of the dipole class. However, Schelkunoff[3] has developed a somewhat different approach to antennas which also leads to answers on input impedance. His method is again based rigorously on Maxwell's equations, but is adapted readily to antennas of many shapes with approximations which are relatively easy to evaluate physically. The method at the same time does an excellent job of filling the need for a physical picture useful for qualitative thinking, a desired goal mentioned earlier. For this reason, we hope over the extent of this chapter to present his approach as a most important modern viewpoint toward antenna theory. Since it does not yet replace older methods for calculation of energy radiated or directivity patterns, but is rather complementary to these, the several useful methods for radiation calculation will be studied with their relation to this viewpoint.

There remains the question of field gradient in corona problems. It is not yet easy to calculate this generally, although approximate results for several antenna shapes should be obtainable from the viewpoint of Schelkunoff, and quite accurate results for spheroidal antennas should follow from the differential equation approach of Stratton and Chu.

[1] E. Siegel and J. Labus, *Hochfrequenz. und Electroakustik*, 43, 166–172 (1934). J. Labus, *Hochfrequenz. und Electroakustik*, 41, 17–23 (1933).

[2] L. J. Chu and J. A. Stratton, *Journ. Appl. Phys.*, pp. 241–248, March, 1941.

[3] S. A. Schelkunoff, *Proc. I.R.E.*, pp. 493–521, September, 1941. S. A. Schelkunoff and C. B. Feldman, *Proc. I.R.E.*, pp. 511–518, November, 1942.

WAVE CONCEPTS OF RADIATION

11.02 The Nature of an Exact Solution to a Radiation Problem

The stated problems of radiation engineering cover a multitude of characteristics of electromagnetic waves. Some of the phenomena, near or directly tied to the conductor, are much like the guided waves of previous chapters. Others, far from the radiator, are like the simple plane waves of Chapter 7. Without going farther, certain aspects of radiation problems may be granted immediately from the experience so far accumulated with Maxwell's equations and electromagnetic fields.

For one thing, the radiator or antenna is a configuration of boundaries (generally metallic and highly conducting, although the use of dielectrics for some boundary formations is quite possible) to which a driving force, such as an electric field between two points, can be applied. This force, speaking in terms of the solution by Maxwell's equations, completes the definition of boundary conditions around the antenna. Currents, charges, and fields then appear around the source, their relative magnitudes and distributions determined by the antenna configuration. Electromagnetic waves must also appear in the region surrounding the source and the antenna to infinity unless the local system is completely surrounded by perfect conductors. Strictly, all boundary conditions about the antenna should be considered, such as the earth and neighboring antennas, excited or unexcited, but the extension in concept is evident, although it may be impossible mathematically and sometimes unimportant practically.

To make these concepts concrete, consider a specific example. Figure 11.02 shows two conducting cones immersed in space. These are coaxial and with only an infinitesimal gap between their apices, A and B. The voltage (the integral of electric field) is specified across the gap between A and B. The cones end at $r = l$. The case will be recognized as an antenna of the so-called dipole type, of special shape. It may be referred to as the *biconical antenna*. The thinking, with certain revisions, will apply to antennas of more common shape. All boundary conditions for the problem are given, and if we could pick out the mathematical functions that satisfy Maxwell's equations and fit these boundaries, all the questions listed in Art. 11.01 could be answered.

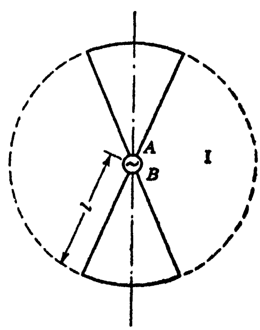


FIG. 11.02. Axial section through biconical antenna.

Now this method of solving for everything at once, fields, currents,

charges, and input impedance, is well-nigh impossible mathematically except for the very simplest cases. Yet a study of the method will be invaluable as: (1) a means for obtaining a firm basis for useful approximate methods of calculating certain parts to the whole problem, which, although used for many years without such backing, can now be applied with confidence to more new cases; (2) a means for attacking certain other phases of the problem by suggesting practical quasi-exact methods.

In the problem of Fig. 11.02, the symmetry and form of the conductors suggests that spherical coordinates be used and all space divided into two regions separated by the fictitious sphere shown dotted in the diagram. A series of solutions in spherical coordinates is next written for each region, of a form general enough to describe the fields. The constant coefficients in these series are evaluated through the boundary conditions by a technique similar to that used in Chapter 3, and once the coefficients are obtained, the whole problem is solved. These boundary conditions are: (1) tangential electric and magnetic fields must be continuous across the dotted boundary sphere; (2) the field at infinity must vanish; (3) the tangential electric field must vanish over the cones (assuming perfect conductivity); (4) the field between cones at their tips is as specified by the driving generator.

The method outlined is mathematically an involved one, but with certain approximations it is not an impossible one, and it should emphasize the fact that the radiation problem has no mystery about it; no new effect has been neglected in previous electromagnetic wave studies. The problem of solving Maxwell's equations subject to boundary conditions is the one solved for waves guided in a wave guide, or along a line, or in a cavity resonator. There may be obvious differences such as the complexity of the boundaries in the antenna problem, the region to infinity as one boundary, and the more frequent occurrence of sources in the form of impressed voltages. These are not fundamental differences, but differences of detail.

11.03 The Antenna as a Finite Length Guide with Reflection and Radiation at the End

To continue the discussion of the specific problem of the biconical antenna, it may be recognized that out of the many electromagnetic solutions for the region I in Fig. 11.02, two of the waves which appear are actually the outgoing and returning principal waves of a conical transmission line studied in Art. 9.09. Assuming perfect conductors, such waves have electric field lines perpendicular to the conical surfaces and therefore always satisfy the boundary condition along the cones. Furthermore, as was shown in Art. 9.09, these waves possess a net inte-

gral of electric field between cones, even at the origin (between the cone apices). Propagation of the principal waves is at the velocity of light for the dielectric material between cones, and all transmission line equations may be used with characteristic impedance given by Eq. 9.09(9).

Since these principal waves are capable of satisfying the boundary conditions along the cones, and the boundary condition of applied voltage at their tips specified by the presence of the generator voltage, then they will certainly be present in region I and may be expected to play a leading role in determining the total field, the current, and the charge distribution in that region, unless the remaining boundary condition, the transition from region I to II, decides otherwise. Let us consider this matter further. First recall that in the above principal waves, E lines lie in surfaces of spheres concentric with the cones' tips, as in Fig. 11.03a. If the conical line were to continue to infinity, only the outgoing one of

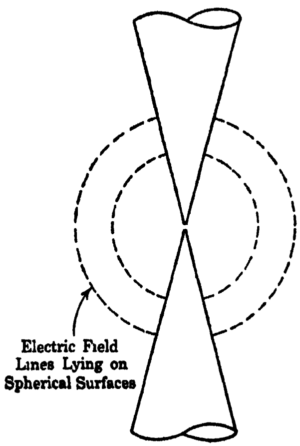


FIG. 11.03a. Portion of conical line.

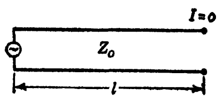


FIG. 11.03b. First approximation to the equivalent circuit of a biconical antenna of length l.

the principal waves would be started by the generator. This wave would by itself satisfy all boundary conditions and no other waves would be needed. Once the cones end, as in the biconical antenna, it may be guessed as a first approximation that current must drop to zero, and that the returning principal wave will be reflected from the end as from the open end of a transmission line. Thus the first approximation to an equivalent circuit would be the open-circuited transmission line of Fig. 11.03b. Current distribution as a function of radius, from this equivalent circuit, would be a simple sinusoid with zero current at the open end. However, when the cones come to the abrupt end, the end effect must be unsatisfied by waves which can only provide fields along the spherical surface. The lack of uniformity which the discontinuity introduces calls for the introduction of other, higher order waves with more complex distributions, in addition to the reflected or returning principal wave.

The situation is not different in many respects from discontinuities

which arise in ordinary guides or transmission lines (Art. 9.15) when a change of cross section occurs. Figure 11.03c shows this in a coaxial transmission system. In region I there must be two principal waves, one traveling towards the change of cross section and one away from it. These two waves in I are not sufficient to satisfy the boundary condition between I and II because they insist upon a certain distribution of radial electric field, one that decreases directly with distance from the axis and is not zero anywhere. But the boundary includes, for example, the section A-B, over which the radial electric field must be zero. Thus, as was discussed in Art. 9.15, there will be local waves started at the boundary in addition to the two principal waves (two of the mathematical functions out of the series needed to completely describe the field). The term local waves is used for the higher order waves here since the dimensions of the cross section in most practical cases cause these waves

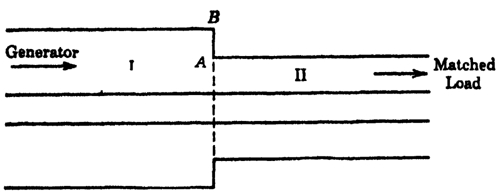


FIG. 11.03c. Discontinuity in coaxial line.

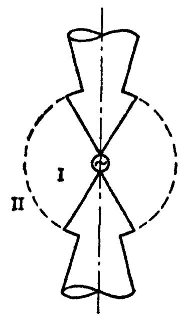


FIG. 11.03d. Discontinuity in conical line.

to be below cut-off. In region II there is a transmitted principal wave and again a series of higher order waves which will be highly localized if the cross-sectional dimensions are small compared with wavelength.

In the conical line case, similar reasoning should apply. Figure 11.03d shows a conical transmission system in which such a change of cross section occurs. What has been said of the waves in regions I and II of the coaxial system of Fig. 11.03c applies, at least in a qualitative way, also to the conical system of Fig. 11.03d. Finally, the biconical antenna of Fig. 11.02 can be looked upon as a wave guiding system immersed in free space, the change of cross section simply being more grandiose than in the transmission line examples. The major difference, speaking qualitatively and generally, is that since the conducting boundaries are absent in region II of the antenna case, then we do not reach into the series of waves in that region to pick out a principal term. the trans-

mitted principal wave of the transmission line type. Instead emphasis is placed upon the more complex functions, the higher order waves that make up the transmitted or radiated waves filling region II.

11.04 The Principal Waves' Currents and Fields as the Radiation Source

As the next step in consideration of the biconical antenna, we could well set down the mathematical functions and compute amplitudes and power flow in the various waves for a number of practical cases. We have, however, reached a point in the discussion where reasonable justification can be given to the most often used approximate methods which are the most practical ones for any but the simplest of antenna structures, and perhaps for these also. We shall accordingly use the wave concepts built up to pass without delay to the approximate methods, returning to the exact approach later in the chapter.

Recall that in the summary of radiation problems, Art. 11.01, it was pointed out that calculations of field patterns and radiated power have most often been made by integration of effects from assumed current distributions about the radiator. From the previous wave study, we have now some basis of knowledge about these current distributions — at least for the biconical antenna. For, in region I of Fig. 11.02, the currents and charges on the conductors and the fields over the spherical boundary can be thought of as contributions from the two principal waves and the series of higher order waves. Now it can be shown that the higher order waves in the biconical antenna have such an amplitude distribution between cones that their effect is relatively inappreciable except very close to the ends of the cones. The current distribution is thus close to sinusoidal, as any combination of the two principal waves would require and as was predicted from the first approximation to an equivalent circuit, Fig. 11.03*b*. As the angle between cones approaches 180°, and the cones degenerate to two thin wires, the current approaches the sinusoidal even more closely. Since fields at any distance from the antenna may be considered as a linear superposition of effects from the important principal wave current and the less important higher order wave currents, an excellent approximation should be the assumption of the principal wave or sinusoidally distributed currents alone as a first step in calculating radiation fields.

When radiation calculations are thus made from the principal wave current distributions only, do not conclude that the higher order waves are completely neglected. For, although they are neglected in region I, they are not neglected in region II. If only the two principal waves were considered, their reflection at the cone endings would lead to a high volt-

age or high E_θ at $r = l$, this field being confined to spherical surfaces. Just outside the region I there would be no field, thus resulting in an impossible discontinuity in tangential electric fields. The fact that fields in region II will be obtained by our calculations shows that this degree of error is not being made.

The approximation can actually be considered as the first step in a converging step-by-step method of which we shall later go to the second step. These steps are:

1. Assume only principal waves in region I.
2. Calculate corresponding higher order waves in region II.
3. Calculate higher order waves in region I to match radial field components of the region II higher order waves obtained in step 2.
4. Correct back and forth through as many succeeding steps as required.

We consequently leave the wave approach for the time being to consider direct integration methods which amount to calculation of step 2 from step 1, and although results will not usually be expressed as a series of waves, they might be, as will be shown later. Step 1 may start from either current or field distribution in the principal wave. The utilization of current distributions will be explored first, then field distributions. In each case power flow will be computed from the Poynting theorem, which must first be proved more generally.

POYNTING CALCULATIONS WITH CURRENTS ASSUMED ON THE ANTENNA

11.05 Poynting's Theorem on Electromagnetic Power Flow in General

In Art. 7.03 the Poynting vector was introduced, which with certain cautions could be interpreted as representing the power flow per unit surface. Now, as a preliminary to computation of radiated power from an antenna, we need to generalize this notion by showing that the integration of $\bar{P}$ over any closed surface will yield the net power flow passing through that surface, and hence the total power leaving the bounded volume, even though sources and dissipations may be included in the volume. These were left out of the early proof of the theorem, all energy flow at that time arising from changes in stored electric and magnetic energy.

Maxwell's equations, written in terms of the total fields, currents, and charges of a region, describe the electromagnetic behavior of the region.

The two curl equations are:

$$\nabla \times \bar{E} = -\frac{\partial \bar{B}}{\partial t} \qquad [1]$$

$$\nabla \times \bar{H} = \bar{\imath} + \frac{\partial \bar{D}}{\partial t} \qquad [2]$$

An equivalence of vector operations, Art. 2.38, shows that

$$\bar{H} \cdot (\nabla \times \bar{E}) - \bar{E} \cdot (\nabla \times \bar{H}) = \nabla \cdot (\bar{E} \times \bar{H}) \qquad [3]$$

If products in (1) and (3) are taken as indicated by this equivalence and added,

$$-\bar{H} \cdot \frac{\partial \bar{B}}{\partial t} - \bar{E} \cdot \frac{\partial \bar{D}}{\partial t} - \bar{E} \cdot \bar{\imath} = \nabla \cdot (\bar{E} \times \bar{H}) \qquad [4]$$

Note that if ϵ is constant,

$$\frac{1}{2} \frac{\partial(\bar{D} \cdot \bar{E})}{\partial t} = \frac{1}{2} \frac{\partial(\epsilon \bar{E}^2)}{\partial t} = \bar{E} \cdot \frac{\partial \bar{D}}{\partial t}$$

Similarly,

$$\frac{1}{2} \frac{\partial(\bar{B} \cdot \bar{H})}{\partial t} = \bar{H} \cdot \frac{\partial \bar{B}}{\partial t}$$

These may be substituted in (4) and all terms integrated over the volume enclosed.

$$\int_V \left[\frac{\partial}{\partial t}\left(\frac{\bar{B} \cdot \bar{H}}{2}\right) + \frac{\partial}{\partial t}\left(\frac{\bar{D} \cdot \bar{E}}{2}\right) + \bar{E} \cdot \bar{\imath}\right] dV = -\int_V \nabla \cdot (\bar{E} \times \bar{H})\, dV$$

From the divergence theorem, Art. 2.12, the volume integral of div $(\bar{E} \times \bar{H})$ must be the same as the surface integral of $\bar{E} \times \bar{H}$ over the surrounding surface.

$$\int_V \left[\frac{\partial}{\partial t}\left(\frac{\bar{B} \cdot \bar{H}}{2}\right) + \frac{\partial}{\partial t}\left(\frac{\bar{D} \cdot \bar{E}}{2}\right) + \bar{E} \cdot \bar{\imath}\right] dV = -\int_S (\bar{E} \times \bar{H}) \cdot d\bar{S} \qquad [5]$$

In this form all terms can be recognized. $\int_V \left(\frac{\bar{D} \cdot \bar{E}}{2}\right) dV$ represents the energy stored in electric fields in the volume. Similarly, $\int_V \left(\frac{\bar{B} \cdot \bar{H}}{2}\right) dV$ represents the energy stored in magnetic fields. The first term must then represent the increase in stored magnetic energy per unit time; the second must represent the increase in stored electric energy per unit time; the third term is the usual ohmic term and so represents energy dissi-

pated in heat per unit time. (Or, if $\bar{\imath}$ is made up of a motion of free charges, $\rho\bar{v}_\rho$, $\bar{E} \cdot \rho\bar{v}_\rho$ represents the energy of acceleration given these charges; and if there are sources, $\bar{E} \cdot \bar{\imath}$ for these sources is of opposite sign and will represent energy added by them.) All the net energy term must have been supplied externally. Thus the term on the right represents the energy flow into the volume per unit time. Changing sign, the rate of energy flow out through the enclosing surface is

$$W = \int_S \bar{P} \cdot d\bar{S} \qquad [6]$$

where

$$\bar{P} = \bar{E} \times \bar{H} \qquad [7]$$

Although it is known from the proof only that total energy flow out of a region per unit time is given by the total surface integral (6), it is often convenient to think of the vector $\bar{P}$ defined by (7) as the vector giving direction and magnitude of energy flow at any point in space. Though this step does not follow strictly, it will not lead us into any pitfalls for our applications.

With Poynting's theorem, we are in a position to calculate energy flow outward from any radiating system for which electric and magnetic fields can be found. Before making such actual calculations, some of the broad concepts which follow from the theorem may first be considered.

The theorem, stating that rate of energy flow out through any closed surface is given by the surface integral (6), shows at once that there can be no energy flow (and hence no radiation) from any system enclosed by a surface over which $\bar{E}$ is everywhere zero, over which $\bar{H}$ is everywhere zero, or over which $\bar{P}$ is directed everywhere tangential to this surface.

Thus for a static charge, there is electric field but no magnetic field, hence no Poynting flux and no energy flow outward. The creation of the charge is another matter, for at the instant of creation, the changing electric effects can cause magnetic field so that the Poynting integral might give a value of energy flow outward.

Similarly, for a steady state direct current flowing in a perfectly conducting system, there is magnetic field, but no electric field so that again there is no Poynting flux and no energy flow outward. The transient involved in building up this current would yield a value, however, just as for the electric charge. If the current is alternating, there is a value of $\bar{E}$ induced by changing magnetic effects. Thus there could be an instantaneous value of energy flow out as calculated from the Poynting theorem at any instant. However, if frequency is low enough so that

times required for propagation of effects about the circuit are truly negligible compared with the period of changing current, this would all return at a later time as energy flow inward. This does not represent radiated energy according to our usual definitions, since for radiated energy we infer a net or average flow outward.

The third case in which no energy can flow outward through a surface is that for which the Poynting vector is directed everywhere tangential to the surface. This must be true if the region is surrounded by perfect conductors, for, by definition, a perfect conductor is one which can support no tangential component of electric field. Total electric field must be always normal to the surface; hence $\bar{E} \times \bar{H}$ must always lie tangential to the surface, and so no electromagnetic energy can pass through the surface.

Problem 11.05(*a*). Suppose that in a long straight wire of circular cross section a direct current is flowing. There is then an axial electric field throughout the wire to overcome the ohmic resistance and a circumferential magnetic field enclosing the wire. Evaluate the Poynting vector over the wire's surface. Compare the result with the I^2R loss in the conductor and state what significance, if any, should be attached to the result.

Problem 11.05(*b*). In this text we are mainly concerned with evaluating the average rate of flow of electromagnetic energy through a surface when $\bar{E}$ and $\bar{H}$ have magnitudes which vary sinusoidally in time. Under these circumstances, $\bar{E}$ and $\bar{H}$ components which yield occasional instantaneous non-zero values for $\bar{E} \times \bar{H}$ may be 90° out of time phase. Thus the contribution to average $\bar{E} \times \bar{H}$ from those components will be zero over the cycle. Show that the average power flow through a surface may be obtained by integrating $\frac{1}{2}Re(\bar{E} \times \bar{H}^*)$ normal to the surface, where $\bar{E}$ and $\bar{H}$ are understood to be peak values expressed in complex form, $\bar{H}^*$ is the conjugate of $\bar{H}$, and Re denotes *real part of*. Show that the average radial component of the Poynting vector in spherical coordinates is

$$(P_r)_{av} = \tfrac{1}{2}Re(E_\theta H_\phi^* - E_\phi H_\theta^*)$$

11.06 The Differential Antenna

In computing radiated power and the field distributions around an antenna when current distribution is assumed over the surface of the antenna's conductors, the simplest example is that of a linear element so short that current may be considered as uniform over its length. Later, certain more complex antennas can be considered as made up of a large number of such differential antennas with the proper magnitudes and phases of their currents. We shall consider only the case in which the current varies sinusoidally with time. Accordingly let it be expressed by $I_0e^{j\omega t}$ or better yet by its peak value I_0 alone with the factor $e^{j\omega t}$ understood.

The direction of the current element will be selected as the z direction,

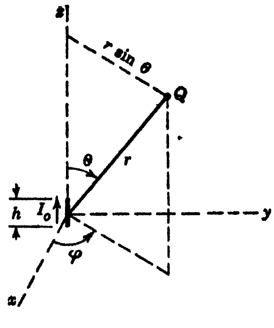


FIG. 11.06. Small current element at origin of spherical coordinates.

and at the origin of a set of spherical coordinates (Fig. 11.06). Its length is h, and it is understood that h is very small compared with wavelength.

Now one way of finding fields once current is given is through the retarded potentials studied in Chapter 4. Article 4.25 gives a form of $\bar{A}$ suitable for present purposes. Since current vector points in the z direction, the vector potential can be only in the z direction. For any point Q at radius r, $\bar{A}$ of Art. 4.25 becomes simply

$$A_z = \frac{hI_0}{4\pi r} e^{-j\frac{\omega r}{v}} \tag{1}$$

Or, in the system of spherical coordinates,

$$\begin{aligned} A_r &= A_z \cos\theta = \frac{hI_0}{4\pi r} e^{-jkr} \cos\theta \\ A_\theta &= -A_z \sin\theta = -\frac{hI_0}{4\pi r} e^{-jkr} \sin\theta \end{aligned} \tag{2}$$

where $k = \omega/v = \omega\sqrt{\mu\epsilon} = 2\pi/\lambda$. There is no ϕ component of $\bar{A}$, and there are no variations with ϕ in any expressions because of the symmetry of the structure about the axis. The electric and magnetic field components may be found directly from the components of $\bar{A}$ by use of the other equations listed in Art. 4.25. Thus,

$$\begin{aligned} H_\phi &= \frac{I_0 h}{4\pi} e^{-jkr}\left[\frac{jk}{r} + \frac{1}{r^2}\right]\sin\theta \\ E_r &= \frac{I_0 h}{4\pi} e^{-jkr}\left[\frac{2\eta}{r^2} + \frac{2}{j\omega\epsilon r^3}\right]\cos\theta \\ E_\theta &= \frac{I_0 h}{4\pi} e^{-jkr}\left[\frac{j\omega\mu}{r} + \frac{1}{j\omega\epsilon r^3} + \frac{\eta}{r^2}\right]\sin\theta \end{aligned} \tag{3}$$

To evaluate average energy flow through a surrounding surface, we may take any surface we please. In particular, if the surface is very far removed from the source, certain terms in the above field expressions are negligible compared with others. These might have been dropped at once, but they were purposely left in to point out certain characteristics of the various components. For the region very near the element (r small) the most important term in H_ϕ is that varying as $1/r^2$. The

important terms in E_r and E_θ are those varying as $1/r^3$. Thus in this region near the element, magnetic field is very nearly in phase with current and H_ϕ may be identified as the usual induction field obtained from Ampère's law. Electric field in this region may be identified with that calculated for the electrostatic dipole, Art. 2.32. (If current is flowing only in one linear direction, a positive charge must accumulate at one end, a negative at the other, thus explaining the dipole solution.) The important components of electric and magnetic field in this region are 90° out of time phase so that these components represent no time average energy flow according to the Poynting theorem.

At very great distances from the source, the only terms important in the expressions for E and H are those varying as $1/r$.

$$
\begin{aligned}
H_\phi &= \frac{jkI_0h}{4\pi r}\sin\theta\, e^{-jkr} \\
E_\theta &= \frac{j\omega\mu I_0 h}{4\pi r}\sin\theta\, e^{-jkr} = \eta H_\phi
\end{aligned}
\qquad [4]
$$

where $\eta = \sqrt{\mu/\epsilon}$.

At great distances from the source, any portion of a spherical wave surface is essentially a plane wave, so the above characteristics typical of uniform plane waves might be expected. E_θ and H_ϕ are in time phase, related by η, and at right angles to each other and the direction of propagation. The Poynting vector is then completely in the radial direction. The time average flow of energy is of interest. The time average of the products of any two sinusoids of equal frequency and of the same phase is one-half the product of their magnitudes. So time average P_r,

$$
P_r = \frac{\eta k^2 I_0^2 h^2}{32\pi^2 r^2}\sin^2\theta \text{ watts/meter}^2
$$

The total energy flow out must be the total surface integral of the Poynting vector over any surrounding surface. For simplicity this surface may be taken as a sphere of radius r. From Fig. 11.06,

$$
\begin{aligned}
W_{av} &= \int_S \bar{P}\cdot \overline{dS} = \int_0^\pi P_r 2\pi r^2 \sin\theta\, d\theta \\
&= \frac{\eta k^2 I_0^2 h^2}{16\pi}\int_0^\pi \sin^3\theta\, d\theta \\
&= \frac{\eta\pi I_0^2}{3}\left(\frac{h}{\lambda}\right)^2 \text{ watts}
\end{aligned}
\qquad [5]
$$

For

$$\eta = \eta_0 = 120\pi \quad \text{ohms}$$

$$W = 40\pi^2 I_0^2 \left(\frac{h}{\lambda}\right)^2 \quad \text{watts} \qquad [6]$$

A radiation resistance may be defined as the resistance which would dissipate the same amount of power with this same constant current flowing.

$$W = I_{\text{rms}}^2 R_r = \frac{I_0^2 R_r}{2}$$

So

$$R_r = 80\pi^2 \left(\frac{h}{\lambda}\right)^2 \quad \text{ohms} \qquad [7]$$

11.07 The Long Straight Antenna

If the antenna length is appreciable compared with wavelength, which is true of practical antennas, current may not be considered constant over the length. This is readily seen in the reasoning of Arts. 11.02 to 11.04. The antenna can, however, be broken into a large number of the differential elements of the type analyzed in Art. 11.06 and the fields from all of these superposed. Although fields or potentials, which are proportional to current, may be superposed, power, which varies as square of current, may not. Thus to use the integration method employed in Art. 11.06, which we shall call the Poynting method, will require that the total $\bar{E}$ and $\bar{H}$ be first evaluated at each point of the large enclosing sphere.

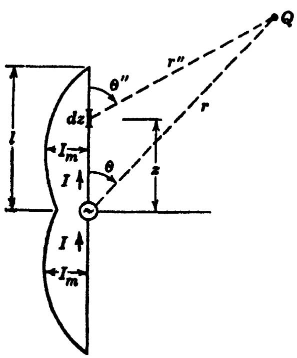


FIG. 11.07. Long straight dipole antenna.

Consider first the long antenna in free space with voltage applied at its midpoint, Fig. 11.07. The resulting antenna is often called a dipole (though admittedly this may give excuse for confusion with the infinitesimal dipole solution of Art. 11.06). For thin wires, the biconical antenna study of Arts. 11.02 and 11.04 applies directly, the cone angles approaching zero. We consequently assume a sinusoidal distribution of current on the antenna with zero current at the ends, following the reasoning of Art. 11.04. This sinusoidal standing wave of current is in time phase over all the antenna.

Then

$$I = \begin{cases} I_m \sin [k(l - z)] & z > 0 \\ I_m \sin [k(l + z)] & z < 0 \end{cases} \qquad [1]$$

From Eq. 11.06(4) the contributions to H_ϕ and E_θ at a great distance r'' from a differential element dz are

$$dE_\theta = \eta dH_\phi = \frac{j\eta kI\,dz}{4\pi r''} e^{-jkr''} \sin\theta''$$

r'' is the distance from any element to Q, whereas r is the distance from the origin to Q. These may be taken so large that the difference between r and r'' is important only as it affects phase, and is completely insignificant in its effect upon magnitude. Similarly, the difference between θ and θ'' will be negligibly small. In the phase difference,

$$r'' = \sqrt{r^2 + z^2 - 2rz\cos\theta} \cong r - z\cos\theta$$

Otherwise,

$$\frac{1}{r''} \cong \frac{1}{r}; \quad \theta'' \cong \theta$$

$$\begin{aligned} E_\theta = \eta H_\phi &= \int_{-l}^{+l} dE_\theta \\ &= \frac{j\eta kI_m}{4\pi r} \sin\theta e^{-jkr} \left\{ \int_{-l}^{0} e^{jkz\cos\theta} \sin [k(l + z)]dz \right. \\ &\qquad \left. + \int_0^l e^{jkz\cos\theta} \sin [k(l - z)]dz \right\} \end{aligned}$$

The integral

$$\int e^{ax} \sin (bx + c)dx = \frac{e^{ax}}{a^2 + b^2} [a \sin (bx + c) - b \cos (bx + c)]$$

So

$$\begin{aligned} E_\theta = \eta H_\phi &= \frac{j\eta kI_m}{4\pi r} \sin\theta e^{-jkr} \left\{ \frac{2}{k \sin^2\theta} [\cos (kl\cos\theta) - \cos kl] \right\} \\ &= \frac{j\eta I_m}{2\pi r} e^{-jkr} \left[\frac{\cos (kl\cos\theta) - \cos kl}{\sin\theta} \right] \end{aligned} \qquad [2]$$

Total $\bar{E}$ and $\bar{H}$ at long distances from the antenna are also at right angles to each other and the direction of propagation, in time phase, and related by η. So, as with the differential antenna of Art. 11.06, the time average

Poynting vector is half the product of field magnitudes.

$$P_r = \tfrac{1}{2}|E_\theta||H_\phi| = \frac{\eta I_m^2}{8\pi^2 r^2}\left[\frac{\cos(kl\cos\theta) - \cos kl}{\sin\theta}\right]^2 \qquad [3]$$

Total power radiated from the long dipole in free space,

$$W = \int_S \bar{P}\cdot d\bar{S} = \int_0^\pi P_r 2\pi r^2 \sin\theta\, d\theta$$

$$= \frac{\eta I_m^2}{4\pi}\int_0^\pi \frac{[\cos(kl\cos\theta) - \cos kl]^2}{\sin\theta}\, d\theta \quad \text{watts} \qquad [4]$$

Since current varies along the antenna, the value of radiation resistance depends upon the current used to define it. Suppose for this case that radiation resistance is defined in terms of maximum current, wherever it may occur.

$$W = \frac{I_m^2 R_r}{2}$$

For

$$\eta = \eta_0 = 120\pi \text{ ohms}$$

$$R_r = 60\int_0^\pi \frac{[\cos(kl\cos\theta) - \cos kl]^2\, d\theta}{\sin\theta} \quad \text{ohms} \qquad [5]$$

It is possible to evaluate the integral in terms of tabulated functions. Graphical integration, however, is often more convenient.

Problem 11.07. The following definite integrals are defined and tabulated.

$$Si(x) = \int_0^x \frac{\sin x}{x}\,dx \quad Ci(x) = -\int_x^\infty \frac{\cos x}{x}\,dx$$

Show that the integration result from (5) may be written

$$R_r = 60\{C + \ln 2kl - Ci(2kl) + \tfrac{1}{2}\sin 2kl[Si(4kl) - 2Si(2kl)]$$
$$+ \tfrac{1}{2}\cos 2kl[C + \ln(kl) + Ci(4kl) - 2Ci(2kl)]\} \text{ ohms}$$

$$C = 0.5772, \text{ Euler's constant}$$

(*Hint:* Substitute $u = \cos\theta$, separate denominator by partial fractions, and note $\lim_{x\to 0} Ci(x) = C + \ln x$.)

11.08 Antennas above Earth

If the earth near an antenna must be taken into account, two very difficult problems can result: (1) effect of earth conductivity, (2) effect of earth curvature. It is common to assume that the earth is plane and perfectly conducting, not alone because it avoids these two difficulties,

but also because it gives answers which agree well with actual results in many practical cases. We shall consequently follow this assumption throughout the chapter when it is necessary to consider the presence of earth.

If earth is assumed plane and perfectly conducting, it is then possible to account for it by imaging the antenna in the earth. For example, given a single cone with axis vertical above earth (Fig. 11.08*a*), the boundary condition of zero electric field tangential to the earth may be satisfied by removing the earth and utilizing a second cone as an image of the first. The problem then reduces to that of the biconical antenna studied previously. Note that current is in the *same vertical* direction at any instant in the two cones. Given a single wire above earth and

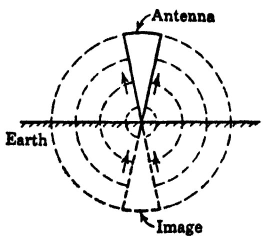


FIG. 11.08*a*. Cone above plane conducting earth and image cone.

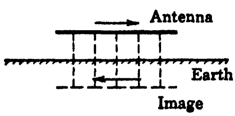


FIG. 11.08*b*. Horizontal wire above plane conducting earth and image wire.

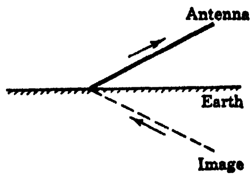


FIG. 11.08*c*. Inclined wire above plane conducting earth and image wire.

parallel to it, as in Fig. 11.08*b*, our knowledge of symmetry in the transmission line problem tells us that the condition of electric field lines normal to the earth is met by removing the earth and placing the image with current in the *opposite horizontal* direction. Generalizing from these two cases, we guess that current direction in the image will be selected so that vertical components are in the same direction, horizontal components in opposite directions at any instant. An example is shown in Fig. 11.08*c*.

The technique of replacing the earth by the antenna image, of course, gives only the proper value of field above the earth plane. The proper value below the perfectly conducting earth plane should be zero. For example, given a long straight vertical antenna above earth, excited at the base, the image reduces the problem to that solved in Art. 11.07. Field strength for maximum current I_m in the antenna is given exactly by Eq. 11.07(2) for all points above the earth $(0 < \theta < \pi/2)$, but is zero for all points below $(\pi/2 < \theta < \pi)$. Thus for power integration, the integral of Eq. 11.07(4) extends only from 0 to $\pi/2$, and radiation resist-

ance is just half that for the corresponding complete dipole

$$W = \frac{\eta I_m^2}{4\pi} \int_0^{\pi/2} \frac{[\cos\,(kl \cos\theta) - \cos kl]^2}{\sin\theta}\, d\theta \quad \text{watts} \qquad [1]$$

Problem 11.08(*a*). Prove by a study of the resulting vector potential the *same vertical direction, opposite horizontal direction* rule for image currents given in the preceding article.

Problem 11.08(*b*). Simpson's rule is useful for evaluation of the radiation integrals. If the area to be evaluated is divided into $2m$ even-numbered portions by $(2m + 1)$ lines spaced an equal distance Δ apart, and values of the function at these lines are $f_0, f_1 \cdots f_{2m+1}$, then the area under the curve is approximately

$$I = \frac{\Delta}{3}\,[(f_0 + f_{2m}) + 4(f_1 + f_3 + \cdots f_{2m-1}) + 2(f_2 + f_4 + \cdots f_{2m-2})]$$

Evaluate the integral of Eq. 11.08(1) for a vertical quarter-wave antenna above earth, $kl = \pi/2$, using $m = 3$ in Simpson's rule. Calculate the radiation resistance and check by the result of Prob. 11.07.

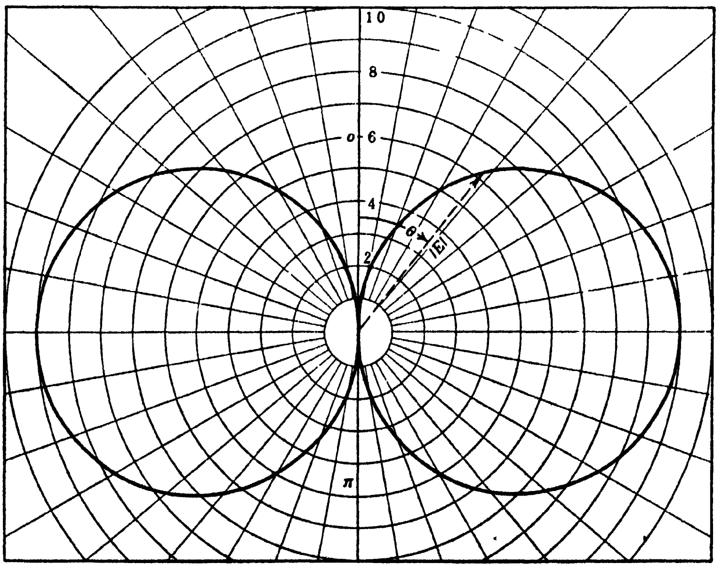


FIG. 11.09*a*. Polar plot of field strength radiation pattern in plane of half-wave dipole.

11.09 Half-Wave Dipole; Quarter-Wave Vertical Antenna above Earth

Article 11.07, in which $l = \lambda/4$, is of special interest to engineers When in free space, the resulting antenna is known as a half-wave dipole By Art. 11.08, results may also be applied directly to a quarter-wave

antenna placed vertically above earth which may be assumed plane and perfectly conducting. The field pattern, Eq. 11.07(2), with $kl = \pi/2$, then reduces to

$$|E_\theta| = \frac{60I_m}{r}\left[\frac{\cos\left(\frac{\pi\cos\theta}{2}\right)}{\sin\theta}\right] \quad \text{volts/meter} \tag{1}$$

and the Poynting vector, Eq. 11.07(3)

$$P_r = \frac{15I_m^2}{\pi r^2}\left[\frac{\cos\left(\frac{\pi}{2}\cos\theta\right)}{\sin\theta}\right]^2 \quad \text{watts/meter}^2 \tag{2}$$

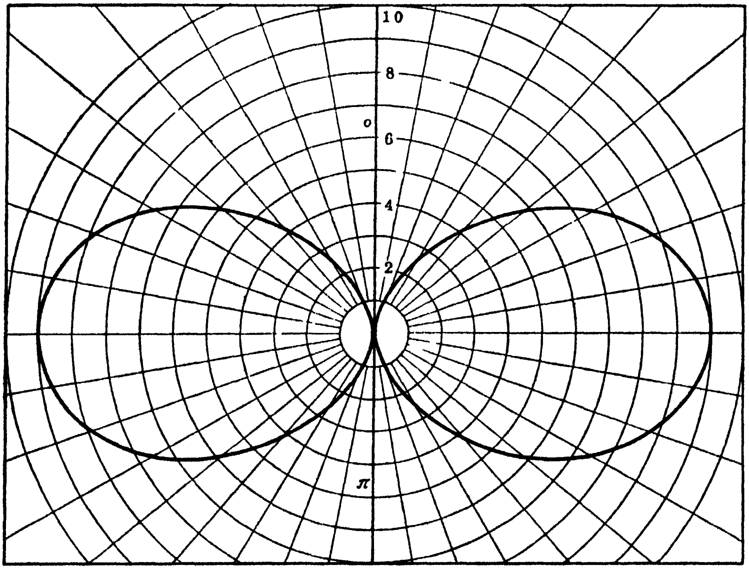


FIG. 11.09b. Polar plot of power density radiation pattern in plane of half-wave dipole.

Polar plots of the bracketed parts of these two functions versus θ are shown in Figs. 11.09a and b. They are useful as field and power or intensity directivity patterns respectively. They apply to the quarter-wave antenna above earth only for $0 < \theta < \pi/2$.

Radiation resistance for the two cases (see Prob. 11.08b) are

Half-wave dipole in space

$$R_r = 73.12 \text{ ohms} \tag{3}$$

Quarter-wave vertical antenna above earth

$$R_r = \tfrac{1}{2}(73.12) = 36.56 \text{ ohms} \tag{4}$$

THE INDUCED EMF METHOD

11.10 Induced EMF Method

At this point we consider an alternative method of calculating power radiated from an assumed current distribution on the antenna. This method is that which arose secondarily out of the discussion of circuit notions in Chapter 5. It consequently seems to hold the possibility of answering some of the questions about antenna input impedance, but although these possibilities are present in concept, they are unfortunately not easily realized practically. The method is of practical use only for the real part of this impedance, or the power radiation term. For the reactive part, the method may be applied conveniently only for a wire of infinitesimal cross section, for which it gives the correct but useless answer of infinite reactance. This situation will be recalled as that found when it was attempted to calculate inductance of a loop of filamentary wire by Neumann's method (Art. 6.15).

If currents are once assumed, then the rigorous field equations give the induced field at every point on the conductor, a part of which is in phase with the current at that point. The product of current density and in-phase induced electric field may then be integrated over the conductor surface to give the total radiated energy. This should give exactly the same answer as integrating the Poynting vector over a remote sphere. The reason why should be apparent if it is considered that:

1. Once currents are assumed over the face of the conductor, the fields outside are fixed and are determined by the same equations in each case.

2. On the surface of the conductor, it is easily seen that since the tangential magnetic field strength is perpendicular to the current density vector and equal to it in magnitude, the product of in-phase induced electric field and current density at the surface is actually that part of the Poynting vector at the surface which has a non-zero time average of outgoing energy. The Poynting theorem in effect states that the time average of the integral must be independent of the enclosing surface of integration, provided, of course, that there are no sources outside of the innermost surface. This last condition is true here since a surface enclosing all the conductor surfaces is certain to include all the currents and charges.

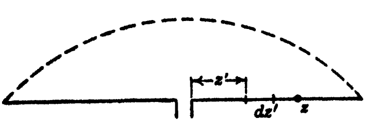


Fig. 11.10.

Since the method has already been used in Chapter 5, there is little more to do but apply it to some more practical cases. Let us develop a form useful for long thin dipole antennas (Fig. 11.10). If current on the

dipole has a distribution $I = I_0 f(z)$ for which $f(z)$ is not yet specified, vector potential at any point z is given as

$$A_z = I_0 \int_{-l}^{+l} \frac{f(z')e^{-jk|z-z'|}}{4\pi|z - z'|} dz' \qquad [1]$$

Radiated power is obtained from the component of $\bar{E}$ tangential to the antenna and in phase with current. We are interested then in the real part of E_z. Referring to Art. 4.25, this real part of E_z is found to be

$$Re(E_z) = -j\omega\mu\left[Im(A_z) + \frac{1}{k^2}Im\left(\frac{\partial^2 A_z}{\partial z^2}\right)\right] \qquad [2]$$

The imaginary component of A_z, if the standing wave pattern $f(z')$ is assumed all of the same phase, is

$$Im(A_z) = -\frac{jI_0}{4\pi}\int_{-l}^{+l} \frac{f(z') \sin k|z - z'|}{|z - z'|} dz'$$

This integral might be evaluated for certain forms of $f(z)$, but it is usually easier to expand the sine in series form.

$$Im(A_z) = -\frac{jkI_0}{4\pi}\int_{-l}^{+l} f(z')\left[1 - \frac{k^2}{3!}(z - z')^2 + \frac{k^4}{5!}(z - z')^4 - \cdots\right] dz'$$

If the function $f(z')$ is an even function, as it should be for the symmetrical dipole, the integral from $-l$ to $+l$ of $f(z')$ multiplied by any odd power of z' is zero, and that for $f(z')$ multiplied by any even power of z' is twice the integral from 0 to l. Thus the above integral becomes

$$Im(A_z) = -\frac{2jkI_0}{4\pi}\int_0^l f(z')\left[1 - \frac{k^2}{3!}(z^2 + z'^2) + \frac{k^4}{5!}(z^4 + 6z^2z'^2 + z'^4) + \cdots\right] dz'$$

The integral can now be broken up into a large number of integrals of the form

$$\alpha_n \equiv \int_0^l (z')^n f(z') dz' \qquad [3]$$

These are definite integrals and can be evaluated simply to give a numerical value for any assumed current distribution. So

$$Im(A_z) = -\frac{2jkI_0}{4\pi}\left[\alpha_0 - \frac{k^2}{3!}(z^2\alpha_0 + \alpha_2) + \frac{k^4}{5!}(z^4\alpha_0 + 6z^2\alpha_2 + \alpha_4) + \cdots\right]$$

and

$$Im\left(\frac{\partial^2 A_z}{\partial z^2}\right)$$

$$= -\frac{2jkI_0}{4\pi}\left[-\frac{k^2}{3}\alpha_0 + \frac{k^4}{10}(z^2\alpha_0 + \alpha_2) - \frac{k^6}{168}(z^4\alpha_0 + 6z^2\alpha_2 + \alpha_4) + \cdots\right]$$

Accordingly

$$Re(E_z)$$

$$= -\frac{2\omega\mu kI_0}{4\pi} \times \frac{2}{3}\left[\alpha_0 - \frac{k^2}{10}(z^2\alpha_0 + \alpha_2) + \frac{k^4}{280}(z^4\alpha_0 + 6z^2\alpha_2 + \alpha_4) + \cdots\right]$$

$Re(E_z)$ as calculated is the component of induced field from changing magnetic effects in phase with current. A component of applied field exactly opposite to this must be applied to support the currents. This represents a net transfer of power at any point, and the total average power in terms of peak current and fields is given by the integral over the wire

$$W = \tfrac{1}{2}\int_{-l}^{+l} |I||E_0|dz = -\tfrac{1}{2}\int_{-l}^{+l} |I|Re(E_z)dz$$

Since I and $Re(E_z)$ are both even functions, this integral is also equivalent to twice the integral from 0 to l. Substituting the expressions for I and $Re(E_z)$ and recalling the definition of α_n,

$$W = \frac{\eta k^2 I_0^2}{3\pi}\left[\alpha_0^2 - \frac{k^2}{5}\alpha_0\alpha_2 + \frac{k^4}{140}(\alpha_0\alpha_4 + 3\alpha_2^2) + \cdots\right] \qquad [4]$$

Radiation resistance in terms of I_0 is then

$$R_r = \frac{2W}{I_0^2} = \frac{2\eta k^2}{3\pi}\left[\alpha_0^2 - \frac{k^2}{5}\alpha_0\alpha_2 + \frac{k^4}{140}(\alpha_0\alpha_4 + 3\alpha_2^2) + \cdots\right]$$

For $\eta = \eta_0 = 120\pi$ ohms,

$$R_r = 80k^2\left[\alpha_0^2 - \frac{k^2}{5}\alpha_0\alpha_2 + \frac{k^4}{140}(\alpha_0\alpha_4 + 3\alpha_2^2) + \cdots\right] \qquad [5]$$

Equation (5) may be applied to symmetrical dipoles of any length by calculating the corresponding integrals for α_n. The results can also be used for vertical antennas over plane earth, using l as height of antenna and taking radiation resistance as half that given by (5), following the image rules of Art. 11.08.

11.11 Half-Wave Dipole Solution by Induced EMF Method

As an example of application of the forms developed in Art. 11.10, let us check the results of Art. 11.09 for a dipole antenna with $2l = \lambda/2$. As before, the current distribution is assumed sinusoidal with phase constant $\beta = k = \omega\sqrt{\mu\epsilon}$.

$$f(z) = \cos kz$$

From Eq. 11.10(3),

$$\alpha_0 = \int_0^{\lambda/4} \cos kz'dz' = \frac{1}{k}$$

$$\alpha_2 = \int_0^{\lambda/4} (z')^2 \cos kz'dz' = \frac{[(\pi/2)^2 - 2]}{k^3} \cong \frac{0.467}{k^3}$$

$$\alpha_4 = \int_0^{\lambda/4} (z')^4 \cos kz'dz' = \frac{[(\pi/2)^4 - 12(\pi/2)^2 + 24]}{k^5} \cong \frac{0.479}{k^5}$$

By substituting in Eq. 11.10(5),

$$R_r = 80[1 - 0.0935 + 0.0081 + \cdots] \cong 73.2 \text{ ohms}$$

The result checks that arrived at previously in Art. 11.09, by the Poynting integration, as it should. The numerical work was relatively easy in this example because of the rapid convergence, only three terms being required. For longer antennas, more terms will be needed.

Problem 11.11. Try the induced emf method on a dipole with $2l = \lambda$. Discuss the significance of the result.

11.12 Degree of Approximation in the Induced EMF Method

There are some worries about the induced emf method which might cause us to question its validity even as an approximate method. These stem mainly from the finite value of $Re(E_z)$ calculated at the surface of the perfect conductor, although the conductor requires a total E_z of zero. Now it has, of course, been carefully pointed out that the value of E_z calculated is only that induced field due to currents and charges on the antenna, and, on the surface of a perfect conductor, must be cancelled by an equal and opposite applied field. But the distribution of induced E_z is fixed once current distribution is assumed. Who is to insure that applied field, which may arise in any one of several different ways, shall have the proper distribution to do this cancelling? No one, of course, for physically the problem is just the reverse. The current distribution adjusts itself so that the induced field exactly cancels the applied field where required by the conductors, and only to the extent that the

assumed current distribution is a good approximation to the actual one, is the net integral of applied field found from the assumed distribution a good approximation to the exact value.

As an extreme example, consider the long dipole with applied field confined at the small gap across its center, and essentially zero everywhere else. Then induced field along the antenna conductor, which is total field there, should itself turn out to be zero and have a value only across the gap. From the assumed sinusoidal distribution, the calculated E_z is found to have a value distributed over all the antenna, not merely concentrated at the gap. It is not obvious at first glance, but only from careful study of the matter, that the integral over the antenna will be approximately the same in the two cases, the difference corresponding only to the difference between the assumed current distribution and the exact one.

For final assurance on the degree of approximation, we return to the argument of Art. 11.10, showing that for a given current distribution assumption the result of power radiated must be the same as that calculated by the Poynting integration. That is, the degree of approximation in the two methods is the same for a given assumed current distribution, and both methods would be exact if the exact current distribution were known.

POYNTING CALCULATIONS WITH FIELDS ASSUMED NEAR THE SOURCE

11.13 The Philosophy of the Assumed Fields Technique

If currents and charges are known, or assumed as known, over the conductors of a radiator, then, as has been seen, it is possible to compute the fields everywhere in the surrounding space and consequently the radiated energy. Now certainly if the charges and currents are known everywhere over the boundaries of a region, then the fields at the boundaries are also known. We could rewrite the previous articles, if desired, with the emphasis on fields over the radiator as the assumed quantities. For example, instead of specifying the charge density at some point on the radiator, one could as well specify the normal electric flux density. Instead of stating the current density flow on the surface, one might write down that the tangential magnetic field is of a certain strength. As a matter of fact, in mks units the assumed quantities would be identical in every way. It also seems almost evident that it is not necessary to know these fields exactly at the conductors of the radiator in order to calculate fields elsewhere, but merely on any complete surface near the radiator.

The approach to radiation calculations which is now to be outlined — the computation of fields in space in terms of specified fields on the boundaries of the space — is one that was instigated by early workers in the theory of radiation.[4] Until relatively recently, their methods have been of minor interest to engineers, at least as compared with the technique of starting directly from assumed currents. The reason has simply been that the types of radiators most practical in the radio broadcast range or even at moderately high frequencies have been such that it was more natural to estimate currents and charge densities on the conductors than the fields near the source. This was so because the antenna structure could be looked at approximately as a section of a circuit or an open line. The growing use of the extremely high frequencies has brought into importance radiators such as horns and open ends of wave guides and coaxial lines, or holes in shields (which though physically small are not always negligibly small compared with wavelength). For these cases it is usually easier to make some estimate of the field distribution over the opening than to do the same for the current and charge distribution along the conductors of the radiator.

The assumption of fields in the above examples is simply guided by a better physical picture than the assumption of current distribution. When the currents and charges are mainly enclosed by the structure, attention seems then to go to the field "escaping" through the opening; these fields seem to be the desired or undesired product of all that has gone on previously inside the structure. At times, for example, the field at the opening might be assumed as that from waves approaching the opening and continuing on past the disturbance as though the boundaries were continuous. Of course, the discontinuity in the guiding would actually disturb the outgoing wave, setting up new waves in effect, both inside and outside the structure. But frequently these new waves may not make enough difference in the field across the opening to prevent a reasonably accurate computation of the radiated field. More often, a good approximation to the actual field across the opening, as far as subsequent radiation calculations are concerned, may be obtained by assuming that the waves approaching the opening on the inside of the radiator are perfectly reflected at the discontinuity or by the surface which contains the opening. In any case, it should be clear that, as with the assumed current technique, the assumed boundary fields would lead to absolutely correct expressions for the fields in the bounded regions provided that the assumed fields themselves were completely correct. The boundary fields are to be assumed by one reasonable guess or another,

[4] Love, *Phil. Trans.*, [A]197, 1901.
MacDonald, *Proc. Lond. Math. Soc.*, 10, 1911; *Phil. Trans.*, [A]212, 295 (1912).

and the accuracy in computing the radiated field depends then on the quality of the original estimates.

11.14 The Equivalent Current Sheet Method

Of the several possible ways of formulating the equations for use with assumptions of field distributions near the source, the method of this article is chosen for the initial study because the work is broken into steps, and it is possible to form relatively sound physical pictures for each of these steps. A neater but less revealing mathematical formulation will be given in a later article. The present study takes advantage of the techniques developed in past studies and follows these stages:

1. Once the fields arising from the source are assumed over some known surface, it is possible to replace the currents and charges of the real source by imaginary current sheets over the surface where fields are assumed, these being selected to produce the fields assumed at that surface with the actual source removed. (The kinds and amounts of currents required are discussed later in this article.)

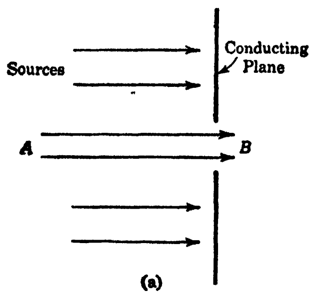


FIG. 11.14a. Conducting plane with aperture.

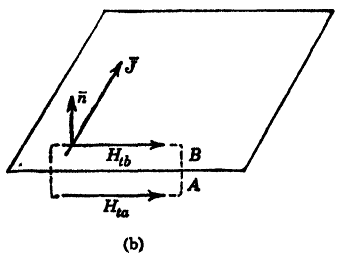


FIG. 11.14b.

2. The problem from this point on is of the same type as that worked previously: from some given distribution of currents, fields may be found and the radiated energy calculated by Poynting's theorem.

Suppose, for an example, that the field to the right of the conducting plane, region B of Fig. 11.14a, is desired. If the exact current and charge sources in the region A to the left of the plane, the sources which actually produce the electromagnetic energy, were known, we could theoretically solve for the desired field, subject to the boundary conditions of the leaky plane. The mathematics would not be pleasant. However, if the field at the surface of the opening arising from those sources were known exactly, it would do just as well, as far as region B is concerned, to replace the actual currents and charges of region A by

fictitious currents and charges lying in the surface of the opening, provided these could be made to produce the same fields at the opening. For, with the conditions over the opening unchanged, and all other boundaries of region B unchanged, the proper solution to Maxwell's equations in region B would be unchanged.

In general, the field that must be produced along the selected surface may have normal and tangential components of electric field, and normal and tangential components of magnetic field. However, at any boundary it is necessary only to know the tangential magnetic and electric fields, for then Maxwell's equations will provide the normal components (Art. 4.23). Let us consider then what fictitious currents must be placed over the opening in the absence of the actual sources to give the same fields in region B that the true sources in region A were causing. The problem is easy and old to us in the case of the tangential magnetic field. For, as indicated in Fig. 11.14*b*, a surface current density J on a given surface will result in a discontinuity of magnetic field components tangential to the surface and normal to $\bar{J}$. That is, the difference between tangential H on one side of the sheet and that on the other,

$$H_{tb} - H_{ta} = J$$

Thus, if the current sheet over the opening in the plane is to replace completely the effect of the sources in A which are producing a given H_t tangent to the boundary, there must be a current density J on this sheet, given by $J = H_t$ in magnitude. The direction will be included if we write

$$\bar{J} = \bar{n} \times \bar{H} \tag{1}$$

where $\bar{n}$ is the unit vector normal to the surface pointing into the region B, and $\bar{H}$ is the total magnetic field. Such a current sheet will wipe out the tangential magnetic field on the A side of the surface (just as though there were no sources in A) but will leave a tangential field as before of magnitude $J = H_t$ on the B side. Thus the current sheet is exactly as effective as the source, which is now assumed absent, in producing tangential magnetic field at the boundary.

Now, the replacing of the sources in so far as they produce tangential electric field would be just as quickly done if only there were such a thing as magnetic currents. Then we could write that the magnetic surface current density $\bar{M}$ is

$$\bar{M} = -\bar{n} \times \bar{E} \tag{2}$$

Also, by analogy with the magnetic vector potential, which for surface

currents is written

$$\bar{A} = \int_S \frac{\bar{J}e^{-jkr}}{4\pi r}\,dS \tag{3}$$

there could be defined an electric vector potential, say,

$$\bar{F} = \int_S \frac{\bar{M}e^{-jkr}}{4\pi r}\,dS \tag{4}$$

Maxwell's equations, if there were such things as magnetic current i_m and magnetic charge ρ_m, would be

$$\begin{aligned} \nabla \cdot \bar{D} &= \rho_e \\ \nabla \cdot \bar{B} &= \rho_m \\ \nabla \times \bar{E} &= -\bar{\imath}_m - \frac{\partial \bar{B}}{\partial t} \\ \nabla \times \bar{H} &= \bar{\imath}_e + \frac{\partial \bar{D}}{\partial t} \end{aligned} \tag{5}$$

and the fields at any distance from the source currents would be obtained from the two vector potentials $\bar{A}$ and $\bar{F}$.

$$\bar{E} = -j\omega\mu\bar{A} + \frac{1}{j\omega\epsilon}\nabla(\nabla \cdot \bar{A}) - \nabla \times \bar{F} \tag{6}$$

$$\bar{H} = -j\omega\epsilon\bar{F} + \frac{1}{j\omega\mu}\nabla(\nabla \cdot \bar{F}) + \nabla \times \bar{A} \tag{7}$$

We have never before written Maxwell's equations to include magnetic currents and charges, so there has never been occasion to integrate them in terms of the augmented group of potential functions, including now the vector electric potential. If anyone does isolate some magnetic charges or demonstrate how to achieve magnetic currents, then the equations just written should take care of the situation. But effects of a magnetic current flow on a surface *can* be realized. We simply cause a tangential electric field along the boundary of a region by some conventional sources on one side of the boundary. Then on the other side of the boundary, an observer, not knowing the nature of the sources, would have as one perfectly logical and possible cause of the tangential electric field some magnetic currents flowing on the surface. If he proceeded on that basis, computing fields everywhere in his region (by the augmented Maxwell equations, of course) he would always get the correct answer.

One misconception may arise at this point. When discussing the general application of this technique to practical problems we have talked much of "guessing" and "estimating" the field over the opening. It may therefore seem that the method of replacing the fields by current sheets, necessarily an approximate one, is in error for just the reason that magnetic current sheets are never actually found in nature. This is entirely incorrect. The approximation comes in any inability, in practice, in predicting the exact field over the complete surface bounding the region. This surface would include, in Fig. 11.14*a*, the plane and infinity in addition to the hole. If the fields were known exactly, no error would be made in replacing tangential fields by surface electric and magnetic currents. As a matter of fact, it is of course unnecessary to bring in the fictitious current sheets at all, for by (1) $\bar{n} \times \bar{H}$ could be written instead of $\bar{J}$ in (3), and by (2) $-\bar{n} \times \bar{E}$ could be written instead of $\bar{M}$ in (4). The two vector functions of space, that yield fields at any point by the differentiations of (6) and (7), would then contain only functions of the fields $\bar{E}$ and $\bar{H}$ over the original surface; the fictitious current sheets with the new concept of a magnetic current would never appear. The completion of such a purely mathematical statement is to be given in Art. 11.16. For the present, if the current sheet concept will be accepted without worry, we see clearly that the calculations to be made are of exactly the same type as those made previously, starting from current distributions along the conductors of the radiator.

Thus, a more detailed breaking down of the steps for this method shows this procedure:

1. Assume intelligently the distribution of electric and magnetic fields over a chosen boundary surface. (This should be over a completely closed surface, and although from the physical reasoning it seems that the closed surface should be one completely surrounding the radiator, the mathematical formulation requires that it be one surrounding the point where fields are to be calculated. Practically, the difference is not important since the region at infinity with zero fields is usually taken as part of the closed surface.)

2. Calculate by (1) and (2) the fictitious electric and magnetic currents that must flow in this boundary if these are to replace completely the actual sources producing the field assumed in step 1.

3. Calculate corresponding magnetic and electric vector potentials, $\bar{A}$ and $\bar{F}$ by (3) and (4).

4. Derive corresponding electric and magnetic fields, $\bar{E}$ and $\bar{H}$, at any desired point by (6) and (7).

5. Calculate radiated power from a Poynting integration.

11.15 Example of Assumed Field Method: Radiation from the Open End of a Coaxial Line

The method of radiation calculations from field assumptions may be applied to a perfectly conducting coaxial transmission line with open end (Fig. 11.15).

If radial dimensions are small compared with wavelength, the wave traveling in the transmission line will be nearly perfectly reflected from the end, and the usual low-frequency concepts of symmetrical tangential magnetic field, radial electric field, and current zero at the open end will be nearly true. For current zero, there is no H_ϕ at the opening, but an $E_{r'}$ alone, given by

$$E_{r'} = \frac{V}{r' \ln (b/a)} = \frac{C}{r'} \qquad [1]$$

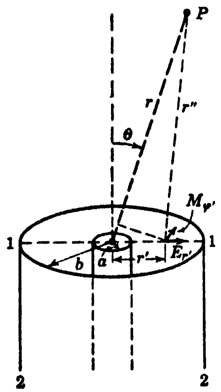


FIG. 11.15. Open end of coaxial line.

An assumption for tangential fields at the opening has now been made, and corresponding equivalent current sheets may be found over this opening. As another portion of the closed surface, select the outside of the transmission line, 1–2. If the line is perfectly conducting, the tangential electric field disappears and so no magnetic current sheet is required. We shall further assume that current flow away from the opening on the outside of the line, like the radiation, is a secondary effect. One effect should not alter the other appreciably, so currents on the outside of the conductor will now be ignored, although from previous chapters we know that such an assumption may be dangerous at some wavelengths. The principal waves inside the line will have equal and opposite currents and no outside magnetic fields will be caused by them. Hence tangential magnetic field will be essentially zero on surface 1–2, and we assume that no electric current sheets are required. The remainder of the enclosing surface is taken at infinity where all fields will have disappeared. There is then only one source, the magnetic current sheet in the opening, given in terms of the assumed electric field of the opening by Eq. 11.14(2). This requires a circumferential magnetic current only.

$$M_\phi = -E_{r'} = -\frac{C}{r'} \qquad [2]$$

For the set of spherical coordinates of Fig. 11.15, the circular symmetry eliminates variations with ϕ. A study of the symmetry shows also

that $\bar{F}$ can have only a ϕ component if $\bar{M}$ has only a circularly symmetric ϕ component. Since there is circular symmetry, any value of ϕ may be chosen for the point P at which field is calculated. The point P will then be chosen in the plane $\phi = 0$. The contribution to F_ϕ at point P from a small element $r'\,d\phi'\,dr'$ at radius r', angle ϕ' on the open transmission line end is then

$$dF_\phi = \frac{M_\phi e^{-jkr''} r'\,dr'\,d\phi'}{4\pi r''} \cos(\phi - \phi') = -\frac{C}{4\pi}\frac{e^{-jkr''}}{r''}\cos\phi'\,dr'\,d\phi'$$

For very large distances from the opening, the difference between r'' and r is important only as it affects phase differences, and this may be found approximately as

$$r'' = \sqrt{r^2 + r'^2 - 2rr'\sin\theta\cos\phi'} \cong r - r'\sin\theta\cos\phi'$$

and

$$F_\phi = -\frac{Ce^{-jkr}}{4\pi r}\int_0^{2\pi}\int_a^b e^{jkr'\sin\theta\cos\phi'}\cos\phi'\,d\phi'\,dr'$$

By the assumption of radial dimensions small compared with wavelength, $kr'\sin\theta\cos\phi'$ is so small that the power series may be substituted for the exponential, with only the first two terms retained.

$$F_\phi = -\frac{C}{4\pi}\frac{e^{-jkr}}{r}\int_0^{2\pi}\int_a^b [1 + jkr'\sin\theta\cos\phi']\cos\phi'\,d\phi'\,dr'$$

$$= -\frac{jkC}{4}\frac{e^{-jkr}}{r}\sin\theta\int_a^b r'\,dr' = \frac{-jkC}{8}\frac{e^{-jkr}}{r}[b^2 - a^2]\sin\theta \quad [3]$$

With $\bar{A} = 0$, no variations with ϕ, and a component of F in the ϕ direction only, the expressions for electric and magnetic fields of Eqs. 11.52(6) and 11.52(7) reduce to

$$\bar{E} = -\frac{\bar{a}_r}{r\sin\theta}\frac{\partial}{\partial\theta}(F_\phi\sin\theta) + \frac{\bar{a}_\theta}{r}\frac{\partial}{\partial r}(r\,F_\phi)$$

$$\bar{H} = -j\omega\epsilon\,\bar{a}_\phi F_\phi$$

So

$$E_r = \frac{jkC}{4}\frac{e^{-jkr}}{r^2}\cos\theta\,[b^2 - a^2]$$

$$E_\theta = -\frac{k^2C}{8}\frac{e^{-jkr}}{r}\sin\theta\,[b^2 - a^2]$$

$$H_\phi = -\frac{\omega\epsilon kC}{8}\frac{e^{-jkr}}{r}\sin\theta\,[b^2 - a^2] = \frac{E_\theta}{\eta}$$

In the radiated field at a long distance from the opening, components varying only as $1/r$ are important. Thus electric and magnetic field, as usual in the radiation field, are at right angles to each other and to radius, are in phase, and are related in magnitude by η. As in the cases previously calculated, the Poynting integration of these fields becomes simply:

$$\begin{aligned} W &= \frac{1}{2}\int_0^\pi |E_\theta||H_\phi| 2\pi r^2 \sin\theta\, d\theta \\ &= \frac{k^4C^2[b^2 - a^2]^2\pi}{64\eta}\int_0^\pi \sin^3\theta\, d\theta \\ &= \frac{k^4C^2[b^2 - a^2]^2\pi}{48\eta} \end{aligned} \qquad [4]$$

$$C = \frac{V}{\ln(b/a)} \qquad \eta = \eta_0 = 120\pi, \quad S = \pi[b^2 - a^2] = \text{Area of opening}$$

So

$$W = \frac{\pi^2V^2}{360}\left[\frac{S}{\lambda^2\ln(b/a)}\right]^2 \quad \text{watts} \qquad [5]$$

Interpreted in terms of an equivalent resistance across the open end,

$$W = \frac{V^2}{2R_r} \quad \text{or} \quad R_r = \frac{180}{\pi^2}\left[\frac{\lambda^2\ln(b/a)}{S}\right]^2 \quad \text{ohms} \qquad [6]$$

There are several questions that may arise in an analysis of this type. These are brought out clearly by this example. First, a value of radiated power is calculated even though the field distribution across the opening was one which neglected the radiation components of field. That is, since it assumed a zero value of a magnetic field H_ϕ, there could be no Poynting vector and no radiated power from the opening. This is exactly analogous to the practice of calculating radiation in conventional antennas from an assumed current which is based upon a distribution neglecting radiation. In both cases these are the major distributions, and slight deviations from this distribution (as must be required if radiation is present) affect the computed field outside very little, especially in the radiation field at a long distance from the source.

The neglect of any currents and charges on the outside of the line may also seem severe, for in spite of the neat little justification, we know that if there is to be any field along the outside, there must also be charges and currents, and the fields calculated in the example will have components there. The result is then not quite correct, for the tangential electric field at least should be zero along the outside of the perfectly con-

ducting line. The results could be improved by finding the currents caused on the line by the first calculation of fields, and correcting for these, repeating until the results are as accurate as desired. However, if it is noted that the field components H_ϕ and E_θ vary as $\sin\theta$, there is little tangential field along this line (θ nearly π) even in the first result, except possibly near the end where θ differs appreciably from π.

Finally, it may seem puzzling that the Poynting integration was made from $\theta = 0$ to π even though the lower half seems shielded by the line. Recall, however, that by the assumption, the conductors of the line, as well as its opening, are replaced completely by the fictitious currents assumed, and so the integration must include all directions.

Problem 11.15(*a*). Repeat the above analysis retaining three terms of the series, of which only two terms were retained in arriving at Eq. 11.13(3). At what diameter to wavelength ratio does the additional term become important?

Problem 11.15(*b*). In the open-ended coaxial example, assume an H_ϕ in the principal wave at the open end of just sufficient magnitude to account for the energy calculated in the first step above. Show that a recalculation of power radiated including the effect of this H_ϕ in addition to the E_r of the first step leads to a truly negligible correction to the first result. Take cross-sectional dimensions small compared with wavelength.

11.16 Other Formulations Utilizing Field Distributions

There are various methods very closely related to that illustrated in Art. 11.15 and which differ in potential accuracy or in formulation. We shall describe these very briefly, intending only that the student be not unaware of important methods for which there is in this text insufficient space. Detailed and excellent treatments of the material which follows will be found in the writings of Schelkunoff[5] and Stratton and Chu.[6]

In the first place, it has already been noted (Art. 11.14) that there is no need for the exact series of steps in which (1) the boundary fields are replaced by current sheets; (2) these are used as integrands in the integrals for the vector potentials; (3) the potentials are differentiated to find the field in space. The entire process can be expressed, by a series of vector manipulations, directly in the form

$$\bar{E}' = -\frac{1}{4\pi}\int_S \{-j\omega\mu(\bar{n}\times\bar{H})\psi + (\bar{n}\times\bar{E})\times\nabla\psi + (\bar{n}\cdot\bar{E})\nabla\psi\}\,dS$$

$$\bar{H}' = \frac{1}{4\pi}\int_S \{-j\omega\epsilon(\bar{n}\times\bar{E})\psi - (\bar{n}\times\bar{H})\times\nabla\psi - (\bar{n}\cdot\bar{H})\nabla\psi\}\,dS \qquad [1]$$

[5] Schelkunoff, *Bell System Tech. Journ.*, **15**, 92 (1936); *Phys. Rev.*, **56**, 308 (1939).
[6] Stratton, "Electromagnetic Theory."
Stratton and Chu, *Phys. Rev.*, **56**, 99 (1939).

$\bar{E}'$ and $\bar{H}'$ are fields at any point inside the surface S, $\bar{E}$ and $\bar{H}$ are the fields on the surface, r is the distance from the differential element dS to the point at which $\bar{E}'$ and $\bar{H}'$ are being evaluated, and $\psi = e^{-jkr}/r$.

In deriving (1) assumptions are made throughout that $\bar{E}$ and $\bar{H}$ are continuous functions. If the example of Art. 11.15 is typical, the assumed $\bar{E}$ and $\bar{H}$ are very likely not to be continuous over the boundary surface. For instance, in Art. 11.15 a magnetic current sheet dropped suddenly to zero in turning the corner from the opening to the remainder of the boundary surface, 1–2 (Fig. 11.15). Certainly, quite apart from assuming fields over the opening which may conceivably be badly off in many cases, it is not pleasant to have calculation formulas which reject the discontinuous functions at the outset. The difficulty, granted, stems from allowing ourselves the liberty of assuming fields over the surface, thus allowing the possibility of choosing discontinuous ones in addition to what might otherwise be simply termed inaccurate ones.

Now it may be that the assumption of a discontinuous field is as close as we can come to estimating the field. This assumption in itself may not be a bad one. Let us not add to its inadequacies, however, by use of formulas which by their derivation should not be applied to discontinuous functions. Maxwell's equations, directly integrated to yield (1), state that the surface $\bar{E}$'s and $\bar{H}$'s cannot be discontinuous unless line currents and charges exist along the contour of the discontinuity. Thus we are led to more complete forms to replace (1) which include contour integrals representing line charges collecting at the discontinuities. Thus to the results of (1) should be added contributions $\bar{E}''$ and $\bar{H}''$ as follows, the integrals being taken around the contour of discontinuity in fields.

$$\begin{aligned} \bar{E}'' &= \frac{1}{4\pi j\omega\epsilon} \oint \nabla\psi \bar{H} \cdot \bar{dl} \\ \bar{H}'' &= -\frac{1}{4\pi j\omega\mu} \oint \nabla\psi \bar{E} \cdot \bar{dl} \end{aligned} \qquad [2]$$

With these additional terms, results should be the same as would be obtained by the method of Art. 11.14, for the line charges were included automatically in that formulation since the form of the equations included the continuity condition for electric and magnetic charges. The forms of (1) and (2) may be useful if many calculations are to be made, although the physical picture given by the equivalent currents of the previous step-by-step method is very helpful in first attacks on a problem.

Stratton and Chu have derived (2) as a vector analogue to Kirchhoff's

theorem. Kirchhoff's theorem gives the value of a scalar quantity at a point inside a source-free region closed by a surface over which the quantity is completely known.

$$\phi' = \frac{1}{4\pi}\int_S \left[\left(\frac{\partial\phi}{\partial n}\right)_S \frac{e^{-jkr}}{r} - \phi_S \frac{\partial}{\partial n}\left(\frac{e^{-jkr}}{r}\right)\right] dS$$

where ϕ' is the value of the scalar function at a point inside, distance r from a point on the surface where the value is ϕ_S. This scalar function satisfies the wave equation,

$$\nabla^2\phi = -k^2\phi$$

The theorem is useful in studies of sound radiation, and in electromagnetic problems for which a scalar quantity may be intelligently assumed over a boundary. It might apply for instance to any scalar component of electric and magnetic field or vector potential in rectangular coordinates, but seldom do these lend themselves to intelligent guesses as to their distribution. For openings large compared with wavelength, as in light diffraction studies, results are not so critically dependent upon the boundary field distribution, and the method may be used to advantage.

In addition to these methods, it is possible also to elaborate on the method of Art. 11.14. Fields may be assumed over only a part of the surrounding surface, as at the open end only of the coaxial line example, and corresponding electric and magnetic current sheets may be placed only over this part of the surface. Electric and magnetic fields from these would then be calculated exactly, subject to the remainder of the boundary geometry. For the example of Art. 11.15, there would be the boundary condition that tangential electric field be zero along the outer surface of the perfectly conducting cylinder. Mathematical difficulties usually require that any attempt to do the problem in this manner requires approximations resulting in about the same series of steps outlined in Art. 11.15.

SYSTEMIZATION OF POYNTING CALCULATIONS

11.17 General Formulas Simplified by Common Approximations

In using the Poynting integration for calculation of radiated power from antennas, either from assumed currents or fields at the source, many of the same mathematical approximations are introduced each time the method is employed. Short cuts are soon discovered and are extremely time saving. It will be desirable therefore to place these on a systematic basis. Schelkunoff has given this systemization in the literature.[7]

[7] Schelkunoff, "A General Radiation Formula," *Proc. I.R.E.*, **27**, 660–666 (October, 1939).

In the Poynting method, field is usually calculated at a great distance from the radiator. The following assumptions are then justified.

1. Differences in radius vector to different points of the radiator are absolutely unimportant in their effect on *magnitudes.*

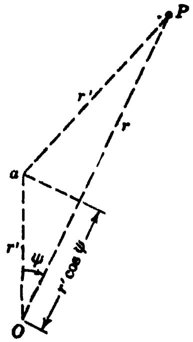


FIG. 11.17.

2. Differences in direction of the radius vector to different points on the radiator are negligible.

3. All field components decreasing with distance faster than $1/r$ are completely negligible compared with those decreasing as $1/r$.

4. Differences in radius vector to different points on the radiator for purposes of finding phase differences are taken as $r' \cos \psi$ of Fig. 11.17, where r' is the radius to the radiating element from the origin, ψ the angle between r' and r, and r is the radius from the origin to the distant point at which field is to be calculated.

Consider the vector potential at point P, distance r from the origin of a radiating system made up of current elements arranged in any manner whatsoever, the element a shown at radius r' from the origin being one of these.

$$\bar{A} = \int_V \frac{\bar{\imath}_a e^{j\omega\left(t - \frac{r''}{v}\right)}}{4\pi r''} dV'$$

By the assumptions listed above, and $e^{j\omega t}$ understood,

$$\bar{A} = \frac{e^{-jkr}}{4\pi r} \int_V \bar{\imath}_a e^{jkr' \cos \psi} \, dV' \qquad [1]$$

$$k = \omega\sqrt{\mu\epsilon} = \frac{\omega}{v}$$

$$\cos \psi = \cos \theta \cos \theta' + \sin \theta \sin \theta' \cos (\phi - \phi') \qquad [2]$$

where θ, ϕ and θ', ϕ' are the angles in spherical coordinates of points P and a respectively. The function of r is now completely outside the integral; the integral itself is only a function of the antenna configuration, current distribution assumption, and direction in which field is to be calculated. Define this integral as the radiation vector $\bar{N}$.

$$\bar{N} = \int_V \bar{\imath}_a e^{jkr' \cos \psi} \, dV' \qquad [3]$$

Then

$$\bar{A} = \frac{e^{-jkr}}{4\pi r} \bar{N} \qquad [4]$$

In the most general case, $\bar{A}$, and hence $\bar{N}$, may have components in any direction. In spherical coordinates, employing the unit vectors,

$$\bar{A} = \frac{e^{-jkr}}{4\pi r}[\bar{a}_r N_r + \bar{a}_\theta N_\theta + \bar{a}_\phi N_\phi]$$

A study of the equation $\bar{H} = \nabla \times \bar{A}$ in spherical coordinates (Art. 2.38) shows that the only components which do not decrease faster than $1/r$ are:

$$H_\theta = -\frac{1}{r}\frac{\partial}{\partial r}(rA_\phi) = \frac{jk}{4\pi r}e^{-jkr}N_\phi$$

$$H_\phi = \frac{1}{r}\frac{\partial}{\partial r}(rA_\theta) = -\frac{jk}{4\pi r}e^{-jkr}N_\theta \qquad [5]$$

An examination of

$$\bar{E} = \frac{1}{j\omega\epsilon}\nabla(\nabla\cdot\bar{A}) - j\omega\mu\bar{A}$$

shows that the only components of $\bar{E}$ which do not decrease faster than $1/r$ are:

$$E_\theta = -\frac{j\omega\mu}{4\pi r}e^{-jkr}N_\theta \qquad E_\phi = -\frac{j\omega\mu}{4\pi r}e^{-jkr}N_\phi \qquad [6]$$

Note

$$\omega\sqrt{\mu\epsilon} = k = \frac{2\pi}{\lambda} \qquad \sqrt{\frac{\mu}{\epsilon}} = \eta$$

So

$$\frac{\omega\mu}{4\pi} = \frac{\eta}{2\lambda}$$

Since (5) and (6) show that $\bar{E}$ and $\bar{H}$ are in time phase and at right angles in space, the Poynting vector

$$\bar{P} = \bar{E}\times\bar{H}$$

has a time average value

$$P_r = \frac{1}{2}\times\frac{\eta}{2\lambda r}\times\frac{1}{2\lambda r}[|N_\theta|^2 + |N_\phi|^2] \qquad [7]$$

Total time average power radiated is

$$W = \int_0^\pi\int_0^{2\pi} P_r r^2 \sin\theta\, d\theta\, d\phi = \frac{\eta}{8\lambda^2}\int_0^\pi\int_0^{2\pi}[|N_\theta|^2 + |N_\phi|^2]\sin\theta\, d\theta\, d\phi \qquad [8]$$

The expression is independent of r as it should be.

The Poynting vector $\bar{P}$ gives the actual power density at any point. However, to obtain a quantity which gives the information of direction only, define K, radiation intensity, as the power radiated in a given direction per unit solid angle. This is the average value of P on a sphere of unit radius.

$$K = \frac{\eta}{8\lambda^2}[|N_\theta|^2 + |N_\phi|^2] \quad [9]$$

and

$$W = \int_0^\pi \int_0^{2\pi} K \sin\theta \, d\theta \, d\phi \quad [10]$$

A plot of K against direction may then define the radiation pattern. It should be recognized that this is a power radiation pattern and not a field strength radiation pattern.

Circularly Symmetric Currents. If all current in some radiating system is circularly symmetric about an axis, this axis may be taken as the axis of a set of spherical coordinates. $\bar{A}$ (hence $\bar{N}$) can have a ϕ component only.

Then

$$K = \frac{\eta}{8\lambda^2}|N_\phi|^2$$

and $\quad [11]$

$$W = 2\pi \int_0^\pi K \sin\theta \, d\theta$$

Currents All in One Direction. If current in a radiating system flows all in one direction, this may be taken as the direction of the axis of a set of spherical coordinates. $\bar{A}$ (hence $\bar{N}$) can have a z component only.

Then

$$N_\phi = 0 \quad N_\theta = -N_z \sin\theta$$

$$K = \frac{\eta}{8\lambda^2}|N_z|^2 \sin^2\theta \quad [12]$$

Useful Relations for Spherical Coordinates. An expression for the angle ψ in terms of the angles θ, ϕ of the point P, and θ', ϕ' of the element a has been given in (2). It sometimes may be desirable to calculate N_θ and N_ϕ from the cartesian components N_x, N_y, N_z,

$$\begin{aligned} N_\theta &= (N_x \cos\phi + N_y \sin\phi)\cos\theta - N_z \sin\theta \\ N_\phi &= -N_x \sin\phi + N_y \cos\phi \end{aligned} \quad [13]$$

Extension to Include Magnetic Currents. Fictitious magnetic as well as electric currents were found useful in calculating radiation from assumed fields, and an electric vector potential $\bar{F}$ was used by analogy with magnetic vector potential $\bar{A}$. A magnetic radiation vector $\bar{L}$ may be related to vector potential $\bar{F}$ as $\bar{N}$ was to $\bar{A}$. Thus, consistent with the assumptions listed previously,

$$\bar{F} = \frac{e^{-jkr}}{4\pi r}\bar{L} \qquad [14]$$

where

$$\bar{L} = \int_V \bar{\imath}_m e^{jkr' \cos\psi}\, dV' \qquad [15]$$

If electric and magnetic field components are now written in the usual way in terms of these two vector potentials, the only components not decreasing faster than $(1/r)$ are

$$E_\theta = -j\frac{e^{-jkr}}{2\lambda r}(\eta N_\theta + L_\phi) \qquad H_\phi = \frac{E_\theta}{\eta} \qquad [16]$$

$$E_\phi = j\frac{e^{-jkr}}{2\lambda r}(-\eta N_\phi + L_\theta) \qquad H_\theta = -\frac{E_\phi}{\eta} \qquad [17]$$

So the radiation intensity,

$$K = \frac{\eta}{8\lambda^2}\left[\left|N_\theta + \frac{L_\phi}{\eta}\right|^2 + \left|N_\phi - \frac{L_\theta}{\eta}\right|^2\right] \qquad [18]$$

11.18 Example: Circular Loop Antenna

It is interesting to check radiation from a circular loop antenna, Fig. 11.18, to compare with the value arrived at in Art. 5.09 by the method since called the induced emf method. Since there is circular symmetry, Eq. 11.17 (11) holds. To compute K we notice that N_ϕ will be the same for any angle ϕ. It will be calculated for an angle in the plane $\phi = 0$. Also $\theta' = \pi/2$ everywhere. Then by Eq. 11.17(2)

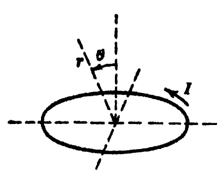


Fig. 11.18. Circular loop carrying current.

$$\cos\psi = \sin\theta\cos\phi'$$

$$N_\phi = I\int_0^{2\pi} e^{jka\sin\theta\cos\phi'}\cos\phi'\, a\, d\phi'$$

$$\cong Ia\int_0^{2\pi}[1 + jka\sin\theta\cos\phi']\cos\phi'\, d\phi'$$

$$= jk\pi Ia^2\sin\theta$$

$$W = \frac{2\pi\eta k^2\pi^2 I^2 a^4}{8\lambda^2}\int_0^\pi \sin^3\theta\, d\theta = \frac{\pi^3\eta k^2 a^4 I^2}{3\lambda^2}$$

For

$$\eta = 120\pi$$

$$W = 10\pi^2 I^2\left(\frac{2\pi a}{\lambda}\right)^4 \quad \text{watts}$$

or

$$R_r = \frac{2W}{I^2} = 20\pi^2\left(\frac{2\pi a}{\lambda}\right)^4 \quad \text{ohms}$$

This value agrees with the result of Art. 5.09.

Problem 11.18. Show that the radiation intensity for a half-wave dipole antenna is

$$K = \frac{15}{\pi} I^2 \frac{\cos^2\left(\frac{\pi}{2}\cos\theta\right)}{\sin^2\theta}$$

11.19 Example: Progressive Wave in Straight Wire of Finite Length

If an unattenuated wave is traveling along a straight wire reaching from $z = 0$ to $z = l$ with velocity $1/\sqrt{\mu\epsilon}$, all the wave being somehow absorbed at the end so that there is none reflected, Fig. 11.19, such a current wave may be represented by $I_0 e^{-jkz'}$. All current is in one direction, so Eq. 11.17(12) may be used.

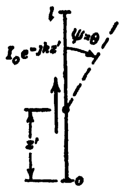


FIG. 11.19. Thin wire of length l supporting a progressive wave.

$$N_z = I_0 \int_0^l e^{-jkz'} e^{jkz'\cos\theta}\, dz'$$

$$= \frac{I_0[1 - e^{-jkl(1-\cos\theta)}]}{jk(1-\cos\theta)}$$

$$|N_z| = \frac{2I_0 \sin\left[\frac{kl}{2}(1-\cos\theta)\right]}{kl(1-\cos\theta)}$$

$$K = \frac{\eta|N_z|^2}{8\lambda^2}\sin^2\theta = \frac{I_0^2\eta}{2\lambda^2}\frac{\sin^2\left[\frac{kl}{2}(1-\cos\theta)\right]}{k^2(1-\cos\theta)^2}\sin^2\theta$$

Also, since there is symmetry about the axis,

$$W = 2\pi\int_0^\pi K\sin\theta\, d\theta = 2\pi\int_0^\pi \frac{I_0^2\eta}{2\lambda^2}\frac{\sin^2\left[\frac{kl}{2}(1-\cos\theta)\right]}{k^2(1-\cos\theta)^2}\sin^3\theta\, d\theta$$

Simplifying and substituting,

$$k = 2\pi/\lambda \quad \eta = 120\pi$$

$$W = 30I_0^2 \int_0^\pi \frac{\sin^3\theta \sin^2\left[\frac{kl}{2}(1 - \cos\theta)\right]}{(1 - \cos\theta)^2} d\theta$$

If the above integral is evaluated[8]

$$W = 30I_0^2 \left[1.415 + \ln\frac{2l}{\lambda} - Ci\frac{4\pi l}{\lambda} + \frac{\sin\frac{4\pi l}{\lambda}}{\frac{4\pi l}{\lambda}} \right] \qquad [1]$$

11.20 Example: Plane Wave Source

The radiation vector and radiation intensity may be calculated for a differential surface element on a uniform plane wave. Such an element might be considered as the elemental radiating source in radiation calculations from field distributions, as was the differential current element for radiation calculations from current distributions.

The plane wave source, that is, one that produces $\bar{E}$ and $\bar{H}$ of constant direction, normal to each other, and in the ratio of magnitudes η over the area of interest may be replaced by equivalent electric and magnetic current sheets over that area, Fig. 11.20.

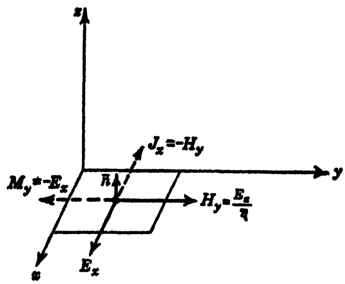


Fig. 11.20. Small plane wave source and equivalent current sheets.

If

$$\bar{E} = \bar{a}_x E_x \quad \text{and} \quad \bar{H} = \bar{a}_y H_y = \bar{a}_y \frac{E_x}{\eta}$$

the equivalent current sheets are

$$J_x = -H_y = -\frac{E_x}{\eta} \qquad M_y = -E_x$$

If this is a source of infinitesimal area dS (actually it need only be small compared with wavelength for following results to hold), the radiation

[8] Stratton, "Electromagnetic Theory," p. 445.

vectors $\bar{N}$ and $\bar{L}$ become simply

$$N_x = -\frac{E_x\,dS}{\eta} \quad L_y = -E_x dS$$

The components in spherical coordinates:

$$N_\theta = -\frac{E_x dS}{\eta}\cos\phi\cos\theta \qquad N_\phi = \frac{E_x dS}{\eta}\sin\phi$$

$$L_\theta = -E_x dS\sin\phi\cos\theta \qquad L_\phi = -E_x dS\cos\phi$$

According to Eq. 11.17(18) the radiation intensity in this case may be given by

$$K = \frac{E_x^2(dS)^2}{\eta 8\lambda^2}[(-\cos\phi\cos\theta - \cos\phi)^2 + (\sin\phi + \sin\phi\cos\theta)^2]$$

$$K = \frac{E_x^2(dS)^2}{2\eta\lambda^2}\cos^4\frac{\theta}{2} \qquad [1]$$

COMBINATIONS OF RADIATORS

11.21 The Superposition of Separate Effects and Inclusion of Mutual Effects

The preceding analyses have involved integrations of effects in space (or over the antenna surface) due to a distribution of assumed currents or fields, so that if there are several complete radiators operating together, currents might be assumed over the entire group, and a complete calculation made for total vector potential $\bar{A}$ (or for the total radiation vectors, $\bar{N}$ and $\bar{L}$). However, as a practical thing, the synthesis of special antennas is often accomplished by putting together elements for which, as isolated antennas, calculations have already been made. Not only may some labor be saved in calculating such cases, but it is also possible in this way to determine the effects of various changes or additions to a structure and thus to attain a desired and special radiation pattern in space. The problem then is to know how to superpose the separate and known radiation characteristics to yield the overall radiation characteristics of the combination.

If calculations for the potentials or fields from each of the radiators operating separately are already available, the fields may be superposed to obtain total fields, and, from these, total power by a Poynting integration. Suppose that in a case with two separate radiators, the component fields are known as $E_{\phi 1}$, $H_{\phi 1}$, etc., due to radiator 1, and $E_{\phi 2}$, $H_{\phi 2}$ due to radiator 2. The Poynting theorem as written in Prob. 11.05(b) then

gives an average value of P_r,

$$P_r = \tfrac{1}{2}Re[\{E_{\theta 1}H_{\phi 1}^* - E_{\phi 1}H_{\theta 1}^*\} + \{E_{\theta 2}H_{\phi 2}^* - E_{\phi 2}H_{\theta 2}^*\} \\ + \{E_{\theta 2}H_{\phi 1}^* + E_{\theta 1}H_{\phi 2}^* - E_{\phi 1}H_{\theta 2}^* - E_{\phi 2}H_{\theta 1}^*\}] \quad [1]$$

The first term is the power due to the first radiator alone; the second is that due to the second radiator alone; the third term is a mutual power due to interaction of fields from the two. The mutual term in the above statement would be obtained in the induced emf method from components of induced fields from charges and changing magnetic effects of the first radiator in phase with currents of the second radiator, and vice versa. It must be emphasized that current distribution assumptions over all radiators are still required for either method, and this may be difficult to make when it is necessary to consider the mutual interaction of several radiators. The induced emf method gives some clue as to this mutual effect upon current distribution, and Carter[9] has by this method calculated mutual effects of parallel linear radiators, interpreting the problem of finding relative current distributions between several such radiators as a circuit problem.

Especially important is the problem of identical radiators with similar current distributions (though magnitudes and phases of currents in individual radiators need not be the same). The radiation vector for one of these alone may be calculated as $\bar{N}_0$. Differences in the distances from the radiators to a far removed point where field is to be calculated is again important only as it affects phase differences, and not as it affects magnitude or direction. Thus the total radiation vector for the system of radiators, if these all have the same orientation, may be written

$$\bar{N} = \bar{N}_0(C_1 e^{jkr_1' \cos\psi_1} + C_2 e^{jkr_2' \cos\psi_2} + \cdots) \quad [2]$$

C_1, C_2, etc., are complex numbers giving the relative magnitudes and phases of currents in the several individual radiators; r_1', r_2', etc., are radii from the common origin to the reference origin of the individual radiators; ψ_1, ψ_2, etc., are the angles between r_1', r_2' and the direction of the radius from the common origin to the point at which field is to be calculated. It follows that the total radiation intensity may be written in terms of the radiation intensity K_0 for one radiator alone.

$$K = K_0 \left| C_1 e^{jkr_1' \cos\psi_1} + C_2 e^{jkr_2' \cos\psi_2} + \cdots \right|^2 \quad [3]$$

The use of these forms will be clearer from the examples to follow.

[9] Carter, *Proc. I.R.E.*, **20**, June, 1932.

Problem 11.21(*a*). A section of parallel wire transmission line when properly terminated may be approximately considered as two wires along which waves are propagating, the currents being opposite in phase at any point along the line. Neglect the radiation from the termination and the mutual effect between the termination and the lines, and compute the radiated power from the line by the method described in Art. 11.21. Use results for a single wire with traveling wave from Art. 11.19.

Problem 11.21(*b*). Given a radiator with horizontal current elements only of radiation intensity K_0, show that if this is placed at a height h above earth which may be assumed plane and perfectly conducting,

$$K = 4K_0 \sin^2 (kh \cos \theta)$$

The vertical direction is taken as the axis, $\theta = 0$.

Problem 11.21(*c*). How is the result of Prob. 11.21(*b*) revised if the $\theta = 0$ axis is taken horizontal and the vertical direction defines $\phi = 0$?

Problem 11.21(*d*). What conclusions similar to those of Prob. 11.21(*b*) can be derived for antennas with vertical current elements only, if placed with their reference origin a distance h above earth?

11.22 Example: Half-Wave Dipoles Separated by a Quarter Wavelength

Consider two half-wave dipoles separated by a quarter wavelength and fed by currents equal in magnitude and 90° out of time phase, as in Fig. 11.22*a*. For a single dipole (Prob. 11.18),

$$K_0 = \frac{15}{\pi} I^2 \frac{\cos^2\left(\dfrac{\pi}{2}\cos\theta\right)}{\sin^2\theta}$$

For the two dipoles with the origin as shown in Fig. 11.22*a*,

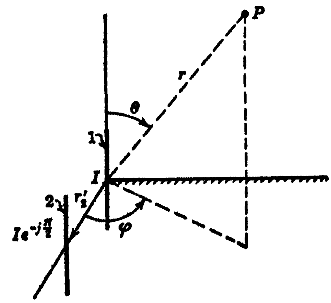


FIG. 11.22*a*. Combination of two half-wave dipoles.

$$r_1' = 0; \quad r_2' = \lambda/4; \quad \theta_2' = \frac{\pi}{2}; \quad \phi_2' = 0$$

So

$$\cos \psi_2' = \sin\theta \cos\phi$$

If

$$I_2 = I_1 e^{-j\frac{\pi}{2}}$$

Then

$$K = K_0 \left|1 + e^{-j\frac{\pi}{2}} e^{j\frac{\pi}{2}(\sin\theta\cos\phi)}\right|^2$$

$$= 4K_0 \cos^2\left[\frac{\pi}{4}(\sin\theta\cos\phi - 1)\right]$$

A horizontal radiation intensity pattern is plotted in Fig. 11.22*b*.

Problem 11.22. Plot radiation intensity patterns for the following half-wave dipole arrays in vertical and horizontal planes.

(*a*) Two parallel dipoles fed in phase with equal currents and placed $\lambda/2$ apart.

(*b*) Four parallel dipoles fed in phase with equal currents and spaced $\lambda/2$ apart.

(*c*) Two dipoles placed end to end, fed in phase with equal currents.

(*d*) Same as (*c*) but with four dipoles.

(*e*) Same as (*d*) but with a perfectly conducting reflecting plane parallel to the dipoles at distance $\lambda/4$.

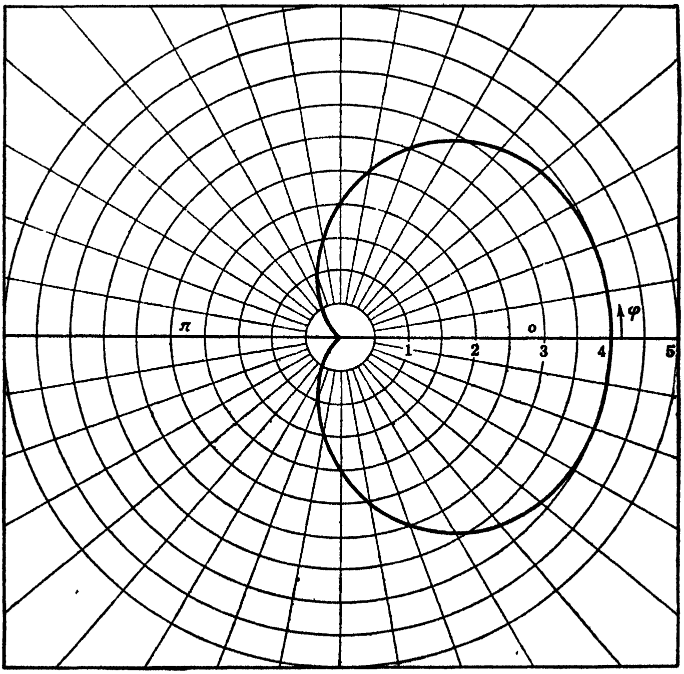


FIG. 11.22*b*. Polar plot of relative power intensity radiation for Fig. 11.22*a* in the plane $\theta = \pi/2$.

11.23 Example: Rhombic Antenna

A rhombic antenna[10] has, oddly enough, the shape of a rhombus, as shown in Fig. 11.23. The antenna is fed at O and terminated at A by the proper resistance so that energy travels along the wires only from O toward A, no reflected waves traveling back from A toward O. This may be analyzed as a system of combined elements, the elements having energy traveling along them in only one direction. Since the elements do not have the same orientation, addition must be by components.

[10] Donald Foster, *Proc. I.R.E.*, **25**, 1327 (October, 1937).
Bruce, Beck, Lowry, *Proc. I.R.E.*, **23**, 24 (January, 1935).

The radiation vector for a single wire with energy traveling at the velocity of light in only one direction, Ie^{-jkl}, has only the direction of the wire. (Art. 11.19.)

$$N_s = \frac{I[1 - e^{-jkl(1-\cos\psi)}]}{jk(1-\cos\psi)} = \frac{I}{jk} f(\psi) \qquad [1]$$

The subscript s denotes the direction of the wire, and ψ is the angle between the wire and the radius vector to the distant point (r, θ, ϕ) at

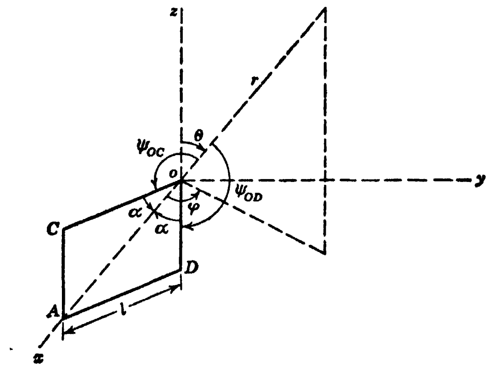


FIG. 11.23. Rhombic antenna and selected coordinate system.

which field is desired. The angles ψ for the various elements, in terms of the coordinates shown in Fig. 11.23, are found to be

$$\begin{aligned} \cos\psi_{OC} &= \cos\psi_{DA} = \sin\theta\cos(\phi+\alpha) \\ \cos\psi_{OD} &= \cos\psi_{CA} = \sin\theta\cos(\phi-\alpha) \end{aligned} \qquad [2]$$

The currents at O for OC and OD are 180° out of phase. They may be taken as I and $-I$. The currents at the beginning of CA and DA (at C and D respectively) are then Ie^{-jkl} and $-Ie^{-jkl}$. Components of radiation vector may now be added, taking into account the differences in phase with respect to the common origin O.

$$r'_{OC} = r'_{OD} = 0 \qquad r'_{CA} = r'_{DA} = l$$

So

$$N_x = \cos\alpha\,[N_1 + N_2 + N_3 e^{jkl\cos\psi_{OC}} + N_4 e^{jkl\cos\psi_{OD}}]$$

$$N_y = \sin\alpha[-N_1 + N_2 + N_3 e^{jkl\cos\psi_{OC}} - N_4 e^{jkl\cos\psi_{OD}}]$$

$$N_z = 0$$

where

$$N_1 = \frac{I}{jk} f(\psi_{OC}) \qquad N_2 = -\frac{I}{jk} f(\psi_{OD})$$

$$N_3 = \frac{Ie^{-jkl}}{jk} f(\psi_{OD}) \quad N_4 = -\frac{I}{jk} e^{-jkl} f(\psi_{OC})$$

and $f(\psi)$ is defined by (1).

Define also

$$S = S(\psi_{OC}\psi_{OD}) = \frac{[1 - e^{-jkl(1-\cos\psi_{OC})}][1 - e^{-jkl(1-\cos\psi_{OD})}]}{(1 - \cos\psi_{OC})(1 - \cos\psi_{OD})} \qquad [3]$$

Then

$$N_x = \frac{IS\cos\alpha}{jk}[\cos\psi_{OC} - \cos\psi_{OD}]$$

$$= -\frac{2IS}{jk}\sin\theta\sin\phi\sin\alpha\cos\alpha \qquad [4]$$

Similarly,

$$N_y = -\frac{2IS}{jk}\sin\alpha(1 - \sin\theta\cos\phi\cos\alpha) \qquad [5]$$

The components of radiation vector in spherical coordinates may be written in terms of the cartesian components N_x and N_y [Eq. 11.17(13)].

$$N_\theta = (N_x\cos\phi + N_y\sin\phi)\cos\theta = -\frac{2IS}{jk}\sin\alpha\sin\phi\cos\theta \qquad [6]$$

$$N_\phi = (-N_x\sin\phi + N_y\cos\phi) = -\frac{2IS}{jk}\sin\alpha[\cos\phi - \sin\theta\cos\alpha] \qquad [7]$$

The radiation intensity is [Eq. 11.17(9)],

$$K = \frac{\eta}{8\lambda^2}[|N_\theta|^2 + |N_\phi|^2]$$

$$= \frac{4I^2\eta}{8\lambda^2k^2}|S|^2[\sin^2\phi\cos^2\theta + (\cos\phi - \sin\theta\cos\alpha)^2]\sin^2\alpha \qquad [8]$$

By trigonometric substitutions in the quantity in brackets,

$$K = \frac{4I^2\eta^2}{8\lambda^2k^2}|S|^2\{[1 - \sin\theta\cos(\phi+\alpha)][1 - \sin\theta\cos(\phi-\alpha)]\}\sin^2\alpha$$

$$= \frac{4I^2\eta}{8\lambda^2k^2}|S|^2[1 - \cos\psi_{OC}][1 - \cos\psi_{OD}]\sin^2\alpha \qquad [9]$$

S, defined by (3), has a magnitude,

$$|S| = \frac{4 \sin\left[\frac{kl}{2}(1 - \cos\psi_{OC})\right] \sin\left[\frac{kl}{2}(1 - \cos\psi_{OD})\right]}{(1 - \cos\psi_{OC})(1 - \cos\psi_{OD})}$$

For air,

$$\eta = 120\pi \quad \text{and} \quad k = \frac{2\pi}{\lambda}$$

so

$$K = \frac{240I^2}{\pi} \sin^2\alpha \frac{\sin^2\left[\frac{kl}{2}(1 - \cos\psi_{OC})\right] \sin^2\left[\frac{kl}{2}(1 - \cos\psi_{OD})\right]}{(1 - \cos\psi_{OC})(1 - \cos\psi_{OD})} \qquad [10]$$

where ψ_{OC} and ψ_{OD} are defined by (2).

From this expression for radiation intensity, it is seen that for large values of $kl/2$, K may become zero many times (each time $\cos\psi_{OC}$, $\cos\psi_{OD}$ are unity, or when $(kl/2)(1 - \cos\psi) = n\pi$). The radiation pattern may then have many lobes. By properly proportioning the angle α and the length l, these lobes may be changed in relative magnitude and the directivity pattern altered greatly.

If a horizontal rhombic antenna is located at height h above a plane earth which may be considered perfectly conducting, the result of Prob. 11.21(b) may be applied directly to find total radiation intensity.

$$K = 4K_0 \sin^2(kh\cos\theta) \qquad [11]$$

K_0 is the radiation intensity for a single rhombic by (10).

Problem 11.23(a). For a rhombic with $l = 3.5\lambda$, $\alpha = 24°$, plot a vertical radiation intensity pattern (in plane $\phi = 0$) and a horizontal pattern (in plane $\theta = \pi/2$).

Problem 11.23(b). How is the vertical pattern revised if the rhombus is placed 2λ above earth?

ANTENNA CHARACTERISTICS BY DIRECT SOLUTION OF THE BOUNDARY PROBLEM

11.24 The Spherical Antenna Problem

Early in the chapter the wave concept of radiation, based upon Maxwell's equations, was studied. By means of this, and particularly by a point of view developed by Schelkunoff, we were able to justify certain assumptions of current or field distributions so that engineering answers could be obtained to the problems of radiation pattern and

total radiated power for many practical antennas. The two methods used, the Poynting method and the induced emf method, did not give information on problems of input impedance, antenna surface gradients, etc., so that for such information we must return to methods based more directly on Maxwell's equations. Of these, the most practical is the method of Schelkunoff referred to previously, but before developing this, the obviously direct and rigorous method discussed in Art. 11.02 will be explored; that is, solution of Maxwell's equations in differential equation form subject to the boundary conditions of the antenna and the applied voltage.

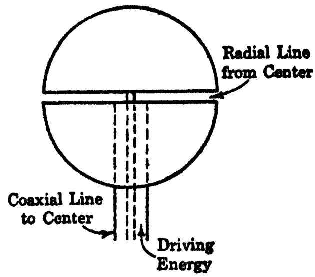


FIG. 11.24. Spherical antenna and possible driving system.

In following articles a simple example will be chosen for which wave solutions can be found in a form suitable for application of the boundary conditions. The specific problem is that of two hemispheres separated by a small gap across which the driving voltage may be applied (Fig. 11.24). In order that this voltage may be applied uniformly about the equator, we might imagine such an antenna driven by a radial transmission line from the center (Art. 9.08), this in turn being connected to a coaxial line which might be brought out at one of the poles in order to disturb boundary conditions very little. This is mentioned only to show that it is practical to excite such an antenna symmetrically, although the presence of the line will be ignored in applying boundary conditions. The hemispheres will be assumed of perfect conductors for the first approximation, losses due to the finite conductivity being obtained as a correction in the approximate manner employed many times previously.

Although results of this problem are possibly of some engineering interest, it is studied mainly for purposes of obtaining background and appreciation of the method in general, and finally of the results of Stratton and Chu who have performed the solution not only for spheres, but also for spheroids. The latter may be considered as excellent approximations to many practical antennas. The wave solutions to be studied will also be those required in the method of Schelkunoff.

11.25 Wave Solutions in Spherical Coordinates

Maxwell's equations in spherical coordinates are written in Art. 4.26. Excellent discussions for complete solutions of these are contained in

an article by Schelkunoff[11] and in Stratton. It is found that solutions to the equations in charge-free dielectrics may be separated into three types, similar to those wave types selected for study of uniform transmission systems (Art. 8.08). Thus an E or transverse magnetic wave may be found with electric field but no magnetic field in the radial direction. An H or transverse electric wave may be found with magnetic field but no electric field in the radial direction. A transverse electromagnetic wave may be found with neither electric nor magnetic field in the radial direction; this last wave type was that principal wave along cones which has been referred to several times in this chapter. For axially symmetric systems, $\partial/\partial\phi = 0$, the field components for the TM and TE wave types separate completely, which is again a similar situation to that found in cylindrical coordinates and rectangular coordinates. The TM wave in this case contains E_r, E_θ, H_ϕ only, and the TE wave contains H_r, H_θ, E_ϕ only. Since the present example, and a large number of engineering problems, are axially symmetric, we shall study only such waves. Also a consideration of the manner of excitation of the spherical antenna, Fig. 11.24, shows that TM waves are definitely excited by the symmetrically applied voltage across the gap, but no reason for any of the TE wave components, H_r, H_θ, or E_ϕ, can be found.

The three curl equations containing E_θ, E_r, and H_ϕ with $\partial/\partial\phi = 0$ are:

$$\frac{\partial}{\partial r}(rE_\theta) - \frac{\partial E_r}{\partial \theta} = -j\omega\mu(rH_\phi) \tag{1}$$

$$\frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(H_\phi \sin\theta) = j\omega\epsilon E_r \tag{2}$$

$$-\frac{\partial}{\partial r}(rH_\phi) = j\omega\epsilon(rE_\theta) \tag{3}$$

Equations (2) and (3) may be differentiated and substituted in (1), leading to an equation in H_ϕ alone.

$$\frac{\partial^2}{\partial r^2}(rH_\phi) + \frac{1}{r^2}\frac{\partial}{\partial\theta}\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}(rH_\phi\sin\theta)\right] + k^2(rH_\phi) = 0 \tag{4}$$

To solve this partial differential equation, we follow the product solution technique. Assume

$$(rH_\phi) = R\Theta \tag{5}$$

R is a function of r alone, Θ is a function of θ alone. If this is substituted in (4), the functions of r may be separated from the functions of θ, and

[11] Schelkunoff, "Transmission Theory of Spherical Waves," *Trans. A.I.E.E.*, 57, 744–750 (1938).

these must then be separately equal to a constant if they are to equal each other for all values of r and θ. For a definitely ulterior motive, we label this constant $n(n + 1)$.

$$\frac{r^2R''}{R} + k^2r^2 = -\frac{1}{\Theta}\frac{d}{d\theta}\left[\frac{1}{\sin\theta}\frac{d}{d\theta}(\Theta\sin\theta)\right] = n(n+1) \qquad [6]$$

Thus there are two ordinary differential equations, one in r only, one in θ only. Let us consider that in θ first, making the substitution

$$u = \cos\theta; \quad \sqrt{1-u^2} = \sin\theta; \quad \frac{d}{d\theta} = -\sin\theta\frac{d}{du}$$

Then

$$(1-u^2)\frac{d^2\Theta}{du^2} - 2u\frac{d\Theta}{du} + \left[n(n+1) - \frac{1}{1-u^2}\right]\Theta = 0 \qquad [7]$$

The differential equation (7) is reminiscent of Legendre's equation (Art. 3.26) and is in fact a standard form. This form is

$$(1-x^2)\frac{d^2y}{dx^2} - 2x\frac{dy}{dx} + \left[n(n+1) - \frac{m^2}{1-x^2}\right]y = 0 \qquad [8]$$

One of the solutions is written

$$y = P_n^m(x)$$

and the function defined by the above solution is called an associated Legendre function of the first kind, order n, degree m. These are actually related to the ordinary Legendre functions by the equation

$$P_n^m(x) = (1-x^2)^{m/2}\frac{d^mP_n(x)}{dx^m} \qquad [9]$$

As a matter of fact, (8) could be derived from the ordinary Legendre equation by this substitution. A solution to (7) may then be written

$$\Theta = P_n^1(u) = P_n^1(\cos\theta) \qquad [10]$$

And from (9)

$$P_n^1(\cos\theta) = -\frac{d}{d\theta}P_n(\cos\theta) \qquad [11]$$

Thus for integral values of n these associated Legendre functions are also polynomials consisting of a finite number of terms. By differentiations according to (9) in Eq. 3.26(8), the polynomials of the first few

orders are found to be

$$
\begin{aligned}
P_0^1 (\cos\theta) &= 0 \\
P_1^1 (\cos\theta) &= \sin\theta \\
P_2^1 (\cos\theta) &= 3\sin\theta\cos\theta \\
P_3^1 (\cos\theta) &= \tfrac{3}{2}\sin\theta\,(5\cos^2\theta - 1) \\
P_4^1 (\cos\theta) &= \tfrac{5}{2}\sin\theta\,(7\cos^3\theta - 3\cos\theta)
\end{aligned}
\qquad [12]
$$

Other properties of these functions that will be useful to us, and which may be found from a study of the above, are:

1. All $P_n^1 (\cos\theta)$ are zero at $\theta = 0$ and $\theta = \pi$.
2. $P_n^1 (\cos\theta)$ are zero at $\theta = \pi/2$ if n is even.
3. $P_n^1 (\cos\theta)$ are a maximum at $\theta = \pi/2$ if n is odd, and the value of this maximum is given by

$$P_n^1(0) = \frac{(-1)^{-\frac{n-1}{2}}\, n!}{2^{n-1}\left[\left(\dfrac{n-1}{2}\right)!\right]^2} \quad n \text{ odd} \qquad [13]$$

4. The associated Legendre functions have orthogonality properties similar to those of sinusoids and Bessel functions studied previously.

$$\int_0^\pi P_l^1 (\cos\theta) P_n^1 (\cos\theta) \sin\theta\, d\theta = 0 \quad l \neq n \qquad [14]$$

$$\int_0^\pi [P_n^1 (\cos\theta)]^2 \sin\theta\, d\theta = \frac{2n(n+1)}{2n+1} \qquad [15]$$

5. The differentiation formula is

$$\frac{d}{d\theta}[P_n^1 (\cos\theta)] = \frac{1}{\sin\theta}[nP_{n+1}^1 (\cos\theta) - (n+1)\cos\theta\, P_n^1 (\cos\theta)] \qquad [16]$$

Note that only one solution for this second order differential equation (7) has been considered. The other solution becomes infinite on the axis, and so will not be required in problems such as the present one where the region of the axis is included in the solution.

To go back to the r differential equation obtainable from (6), substitute the variable $R_1 = R/\sqrt{r}$.

$$\frac{d^2R_1}{dr^2} + \frac{1}{r}\frac{dR_1}{dr} + \left[k^2 - \frac{(n+\frac{1}{2})^2}{r^2}\right]R_1 = 0$$

By comparing with Eq. 3.18(3) it is seen that this is Bessel's differential equation of order $(n + \frac{1}{2})$. A complete solution may then be written

$$R_1 = A_n J_{n+\frac{1}{2}}(kr) + B_n N_{n+\frac{1}{2}}(kr) \qquad [17]$$

and

$$R = \sqrt{r}\, R_1$$

If n is an integer, these half-integral order Bessel functions reduce simply to algebraic combinations of sinusoids.[12] For example, the first few orders are:

$$J_{1/2}(x) = \sqrt{\frac{2}{\pi x}} \sin x \qquad N_{1/2}(x) = -\sqrt{\frac{2}{\pi x}} \cos x$$

$$J_{3/2}(x) = \sqrt{\frac{2}{\pi x}}\left[\frac{\sin x}{x} - \cos x\right] \qquad N_{3/2}(x) = -\sqrt{\frac{2}{\pi x}}\left[\sin x + \frac{\cos x}{x}\right]$$

$$J_{5/2}(x) = \sqrt{\frac{2}{\pi x}}\left[\left(\frac{3}{x^2} - 1\right)\sin x - \frac{3}{x}\cos x\right] \qquad N_{5/2}(x) = -\sqrt{\frac{2}{\pi x}}\left[\frac{3}{x}\sin x + \left(\frac{3}{x^2} - 1\right)\cos x\right] \tag{18}$$

The linear combination of the J and N functions into Hankel functions (Art. 3.16) represent waves traveling radially inward or outward, and boundary conditions will be as found previously for other Bessel functions:

1. If the region of interest includes the origin, $N_{n+\frac{1}{2}}$ cannot be present since it is infinite at $r = \infty$.

2. If the region of interest extends to infinity, the linear combination of J and N into the second Hankel function, $H^{(2)}_{n+\frac{1}{2}} = J_{n+\frac{1}{2}} - jN_{n+\frac{1}{2}}$, must be used to represent a radially outward traveling wave.

The particular combination of $J_{n+\frac{1}{2}}(kr)$ and $N_{n+\frac{1}{2}}(kr)$ required for any problem may be denoted as $Z_{n+\frac{1}{2}}(kr)$ and now by combining correctly (17), (10), and (5), H_ϕ is determined. E_r and E_θ follow from

[12] A very neat notation for the spherical or half-integral order Bessel functions has recently come into use. Thus $j_n(x)$ denotes $\sqrt{\frac{\pi}{2x}}\, J_{n+\frac{1}{2}}(x)$, and similar small letters denote the other spherical Bessel and Hankel functions. (See for example Stratton.) We shall not employ this notation since for our limited use of these functions it will be preferable to retain the usual Bessel function forms for which recurrence formulas and rules of differentiation are listed in Chapter 3

(2) and (3) respectively.

$$H_\phi = \frac{A_n}{\sqrt{r}} P_n^1 (\cos\theta) Z_{n+\frac{1}{2}}(kr)$$

$$E_\theta = \frac{A_n P_n^1 (\cos\theta)}{j\omega\epsilon r^{3/2}} [nZ_{n+\frac{1}{2}}(kr) - krZ_{n-\frac{1}{2}}(kr)] \qquad [19]$$

$$E_r = -\frac{A_n n Z_{n+\frac{1}{2}}(kr)}{j\omega\epsilon r^{3/2} \sin\theta} [\cos\theta\, P_n^1 (\cos\theta) - P_{n+1}^1 (\cos\theta)]$$

These solutions of course represent an infinite number of possible *TM* waves, one for each value of n. For any given problem, as in the present one, it will usually be necessary to add many of these waves to fit the boundary conditions.

Problem 11.25. Obtain the field components E_ϕ, H_θ, H_r for the nth order *TE* wave with $\partial/\partial\phi = 0$.

11.26 Identity of the First Order *TM* Wave with the Dipole Field

Let us digress a moment to look at the first order wave defined by Eq. 11.25(19) with $n = 1$. (Note that for $n = 0$ there is no wave since $P_0^1 (\cos\theta) = 0$.) This will be considered in a region extending to infinity, so that the proper Bessel function is a Hankel function of second kind to represent only a radially outward traveling wave.

$$Z_{3/2}(kr) = H_{3/2}^{(2)}(kr) = J_{3/2}(kr) - jN_{3/2}(kr)$$

From Eq. 11.25(18), this is

$$H_{3/2}^{(2)}(kr) = \sqrt{\frac{2}{\pi kr}}\left[\left(\frac{\sin kr}{kr} - \cos kr\right) + j\left(\sin kr + \frac{\cos kr}{kr}\right)\right]$$

$$= \sqrt{\frac{2}{\pi kr}}\; ^{-jkr}\left(\frac{j}{kr} - 1\right)$$

Thus the three field components reduce to

$$H_\phi = A_1\sqrt{\frac{2k}{\pi}} \sin\theta \left[\frac{j}{k^2r^2} - \frac{1}{kr}\right] e^{-jkr}$$

$$E_\theta = \eta A_1\sqrt{\frac{2k}{\pi}} \sin\theta \left[\frac{j}{k^2r^2} - \frac{1}{kr} + \frac{1}{k^3r^3}\right] e^{-jkr} \qquad [1]$$

$$E_r = \frac{2A_1}{j\omega\epsilon r}\sqrt{\frac{2k}{\pi}} \cos\theta \left[\frac{j}{k^2r^2} - \frac{1}{kr}\right] e^{-jkr}$$

Compare these with Eq. 11.06(3), the field components for a small current element or dipole. They are identical if the constant A_1 is identified as $I_0 hk^{3/2}/4j\sqrt{2\pi}$. It is interesting, although expected, to find that the solution for the infinitesimal dipole is one of the general *TM* wave types. It is also prominent in the spherical antenna solution. Electric field lines for such a wave are sketched in Fig. 11.26.

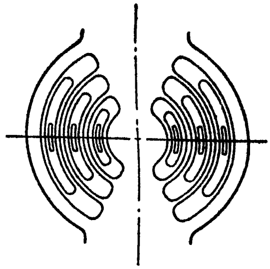


FIG. 11.26. Axial section showing electric field lines in first order symmetrical *TM* waves.

Problem 11.26(*a*). Study the $n = 2$ *TM* wave and show that it corresponds to a quadrupole field, i.e., field from two small current elements at right angles.

Problem 11.26(*b*). Show that the $n = 1$ *TE* wave from Prob. 11.25 corresponds to the field from a small current loop.

11.27 Superposition of *TM* Waves to Match Boundary Conditions of Spherical Antenna

The boundary conditions to be satisfied on the surface of the spherical antenna (assuming perfect conductivity for the sphere in the first approximation) are:

1. $E_\theta = 0$ at $r = a$, except across gap.
2. $E_\theta =$ applied field across gap, $\theta = \dfrac{\pi}{2} - \alpha$ to $\dfrac{\pi}{2} + \alpha$.

The distribution of E_θ across the gap is not known, but its integral across the gap (which is the same as the integral from 0 to π since E_θ is elsewhere zero) must be equal to the applied voltage.

$$V_0 = \int_{\frac{\pi}{2}-\alpha}^{\frac{\pi}{2}+\alpha} E_\theta a \, d\theta = \int_0^\pi E_\theta a \, d\theta \qquad [1]$$

Now if the exact distribution of E_θ across the gap were known, the function E_θ at $r = a$ could be expanded in a series. This series would be written in associated Legendre polynomials so that it might be compared directly with previous *TM* wave solutions. Known functions may be expanded in terms of these functions in a manner similar to that used to expand functions in a Fourier series, a series of Bessel functions, Art. 3.23, or a series of ordinary Legendre polynomials, Art. 3.26. The formula for the coefficients follows from the orthogonality properties of Eqs. 11.25(14) and 11.25(15).

$$f(\theta) = \sum_{n=1}^{\infty} b_n P_n^1 (\cos\theta) \qquad [2]$$

where

$$b_n = \frac{2n+1}{2n(n+1)} \int_0^{\pi} f(\theta) P_n^1 \,(\cos\theta)\sin\theta \, d\theta \tag{3}$$

The exact form of the $f(\theta)$ to be expanded, i.e., E_θ, is not known except that it is zero everywhere but at the gap. If the gap is truly small, we may approximate the answer to the integral (3) by assuming that P_n^1 $(\cos\theta)$ and $\sin\theta$ do not vary appreciably across the gap. That is, assume that P_n^1 $(\cos\theta)$ is approximately constant at its maximum value given by Eq. 11.25(13) and that $\sin\theta$ is constant at its maximum value of unity over the gap. Then,

$$b_n = \frac{2n+1}{2n(n+1)} P_n^1(0) \int_{\frac{\pi}{2}-\alpha}^{\frac{\pi}{2}+\alpha} E_\theta \, d\theta$$

The latter integral may be found directly from (1).

$$b_n = \frac{(2n+1)P_n^1(0)V_0}{2n(n+1)a}$$

and

$$E_\theta\Big|_{r=a} = \sum_{n=1}^{\infty} b_n P_n^1 \,(\cos\theta) \tag{4}$$

The above is exactly correct for an infinitesimal gap, but for any gap of finite size it will not give correct coefficients for the highest harmonics that vary appreciably over the region of the gap.

For the wave solution in the space surrounding the antenna, we are to add an infinite number of the *TM* waves found in Eq. 11.25(19). Following previous reasoning, the Bessel function solution should be the Hankel function of the second kind since the region surrounding the antenna extends to infinity. Then E_θ at $r = a$ from Eq. 11.25(19) may be written

$$E_\theta\Big|_{r=a} = \frac{j}{\omega\epsilon a^{3/2}} \sum_{n+1}^{\infty} A_n P_n^1 \,(\cos\theta)[kaH_{n-\frac{1}{2}}^{(2)}(ka) - nH_{n+\frac{1}{2}}^{(2)}(ka)] \tag{5}$$

By comparing (5) with (4), A_n may be evaluated.

$$A_n = \frac{\omega\epsilon a^{3/2} b_n}{j[kaH_{n-\frac{1}{2}}^{(2)}(ka) - nH_{n+\frac{1}{2}}^{(2)}(ka)]} \tag{6}$$

In (4) b_n is defined, so A_n, the arbitrary coefficients of the solution, are completely determined in terms of applied voltage and the antenna dimensions. Field at any point is now expressed in a series of *TM* waves with determined coefficients. Thus if desired, the field distribu-

tion at any radius could be mapped and thus the radiation pattern obtained. However, we shall go directly to the calculation of antenna impedance. It is at least evident, though, that E_θ and H_ϕ in the radiation field at large distances are zero along the axis and a maximum at $\theta = \pi/2$, since all odd P_n^1 (cos θ) (the only ones excited) are zero at $\theta = 0$, maximum at $\theta = \pi/2$.

11.28 Impedance of Spherical Antenna

The magnetic field H_ϕ is now determined since the coefficients A_n are known by Eqs. 11.27(6) and 11.27(4).

$$H_\phi = \sum_{n=1}^{\infty} \frac{A_n}{r^{1/2}} P_n^1 (\cos\theta) H_{n+\frac{1}{2}}^{(2)}(kr) \tag{1}$$

Surface current density is given in terms of the magnetic field at the conductor surface

$$\bar{J} = \bar{n} \times \bar{H}$$

or

$$J_\theta = -H_\phi\big|_{r=a}$$

Thus total current flow on the antenna at any angle θ is

$$I_\theta = 2\pi a \sin\theta J_\theta = -2\pi a \sin\theta H_\phi\big|_{r=a} \tag{2}$$

The total current flow away from the gap, at $\theta = \pi/2$, from (1) and (2)

$$\begin{aligned} I &= -I_\theta\big|_{\theta=\frac{\pi}{2}} \\ &= 2\pi a \sum_{n=1}^{\infty} \frac{P_n^1(0) A_n}{a^{1/2}} H_{n+\frac{1}{2}}^{(2)}(ka) \end{aligned} \tag{3}$$

A_n as defined by Eqs. 11.27(4) and 11.27(6) is proportional to V_0 so that the ratio of I to V_0 may be written as an admittance.

$$Y = \frac{I}{V_0} = \sum_{n=1}^{\infty} Y_n$$

where [4]

$$Y_n = \frac{j\pi(2n+1)[P_n^1(0)]^2}{n(n+1)\eta} \left[\frac{1}{\dfrac{n}{ka} - \dfrac{H_{n-\frac{1}{2}}^{(2)}(ka)}{H_{n+\frac{1}{2}}^{(2)}(ka)}} \right]$$

As usual $\eta = \sqrt{\mu/\epsilon}$.

The form of (4) is particularly interesting, since it represents total admittance as the sum of a number of admittances, one for each har-

monic solution corresponding to a given n. This is of the form for the admittance of a group of circuits in parallel. Each circuit then corresponds to a given harmonic solution and has admittance characteristics determined by (4). $P_n^1(0)$ is defined by Eq. 11.25(13) and the $H_{n+\frac{1}{2}}^{(2)}$

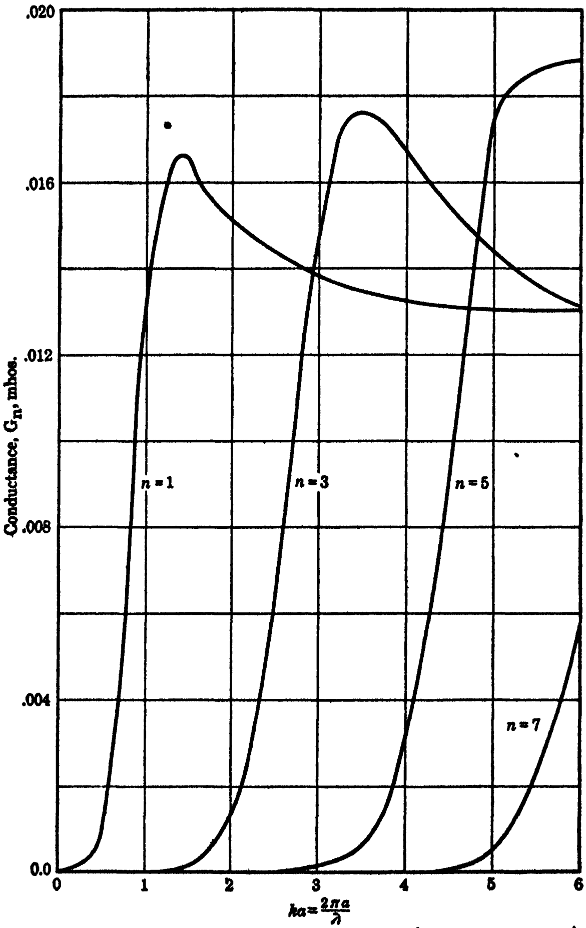


FIG. 11.28*a*. Conductance of individual spherical *TM* wave orders.

functions by Eq. 11.25(18) and the usual definition for Hankel functions, $H_n^{(2)}(x) = J_n(x) - jN_n(x)$. So this admittance characteristic may be calculated. Its conductance and susceptance parts are plotted against ka for air dielectric ($\eta = 120\pi$) in Figs. 11.28*a* and *b*. Note that there are no even harmonics since $P_n^1(0) = 0$ for n even. This is as would be expected since $P_n^1(\cos\theta)$ for n even are all odd functions

with respect to the equator and should not be stimulated by a configuration symmetrical with respect to the equator.

The higher harmonics may be readily approximated:

$$Y_n \cong \frac{j\pi(2n+1)[P_n^1(0)]^2}{\eta n^2(n+1)}\, ka \quad \text{if } ka \ll n$$

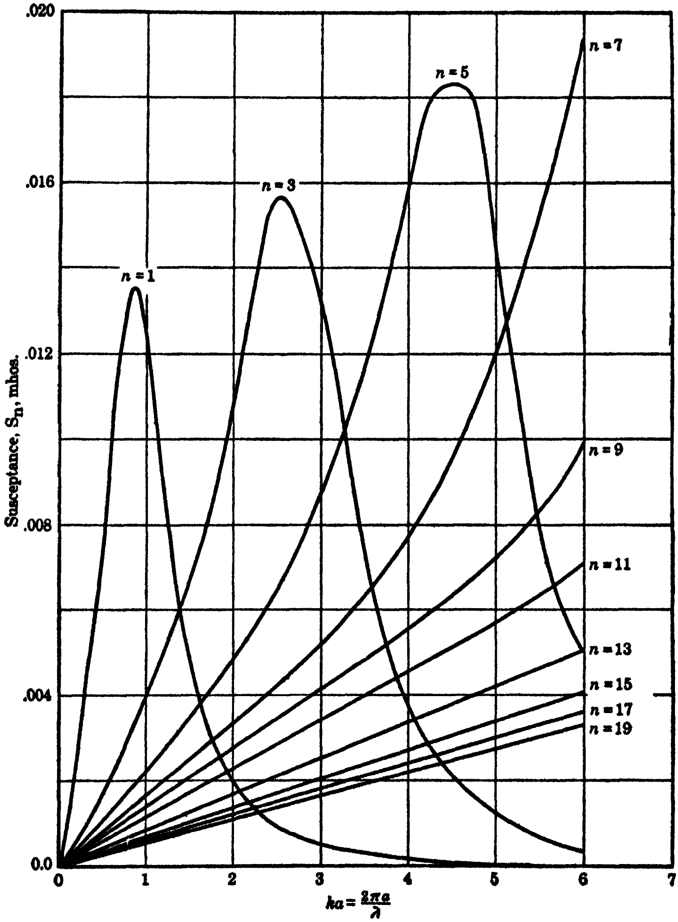


FIG. 11.28b. Susceptance of individual spherical TM wave orders.

A study of this equation will show that if an infinite number of n's are present, the total Y does not converge since finite contributions to susceptance are added by the higher n's forever. However, this is only true for an infinitesimal gap, for which an infinite susceptance term

might be expected. When the gap is finite, a point will be reached at which the coefficients b_n (and hence Y_n) will begin to decrease, approaching zero as n approaches infinity. This occurs for harmonic solutions which vary appreciably in the P_n^1 (cos θ) function over the

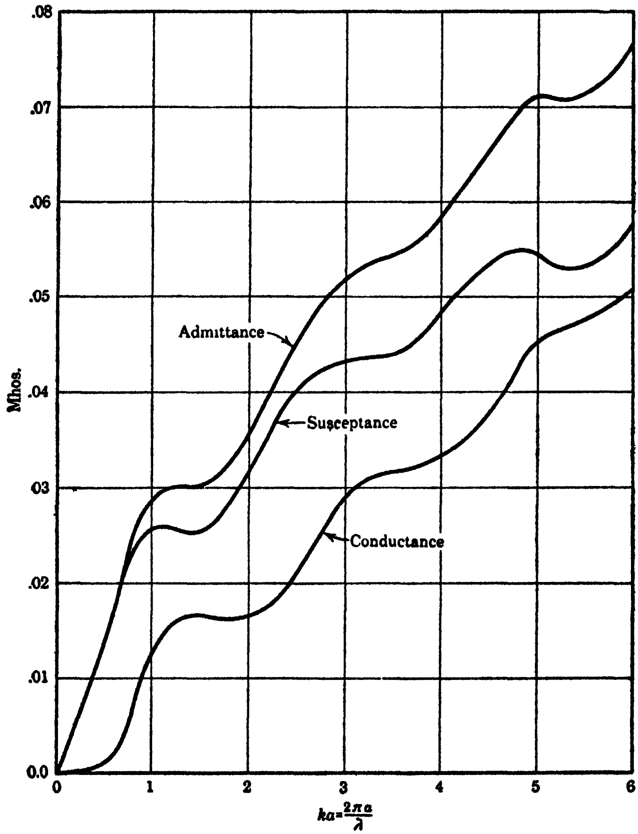


FIG. 11.28c. Total admittance, conductance, and susceptance for spherical antenna.

region of the gap. Consequently, the actual total admittance cannot be obtained until the width of the gap is known. However, the form of the curve and the order of magnitude of admittance will be changed little by missing the point by a few n's above which contributions to Y from Y_n should cease. Consequently, a representative curve for $\eta = 120\pi$ is plotted in Fig. 11.28c, using up to $n = 19$. The conductance or real part does converge and so the curve for conductance should be quite accurate.

These conclusions from the admittance curves are of importance.

1. Admittance of any mode, Y_n, is zero at zero frequency.
2. For low frequencies ($ka \ll n$) admittance is mainly a susceptance proportional to frequency, thus representing a pure capacitance: the capacitance between the hemispheres.
3. In the region of $ka = n$ both conductance and susceptance reach a maximum.
4. Input admittance is capacitive at any frequency.
5. Input conductance at high frequencies approaches a constant value.
6. Total admittance curve has very flat portions in the neighborhood of $ka = 2, 4, 6$, etc.

Problem 11.28(*a*). Show that admittances in a given mode in air approach these values at low and high frequencies.

$$Y_n \to jkaK_n \quad ka \ll n$$

$$Y_n \to K_n \quad ka \gg n$$

where

$$K_n = \frac{(2n+1)[P_n^1(0)]^2}{120n(n+1)} \quad \text{mhos}$$

Problem 11.28(*b*). Calculate voltage required to radiate 100 watts at the first flat point on the admittance-frequency curve, Fig. 11.28*b*.

Problem 11.28(*c*). Find the point of maximum gradient E_r in the antenna, and calculate approximately its value in terms of applied voltage. Take ka in the vicinity of unity. (*Suggestion:* calculate only that in the predominant wave mode.)

11.29 Radiation Efficiency of Spherical Antenna

The spherical antenna has been assumed to be perfectly conducting, but the first order correction for finite conductivity may be made in the usual way by finding the losses due to the current flow calculated for the ideal conductors, but flowing in the sphere of known conductivity. Moreover, to find order of magnitude of these losses, it will be noted that at a particular frequency the major part of this loss arises from current in the mode which predominates at that frequency. For example, in the neighborhood of $ka = 1$, current is mainly in the $n = 1$ mode. Current density J_θ is given by H_ϕ, Eq. 11.28(1). This may be written in terms of the admittance defined in Eq. 11.28(4). For the nth mode,

$$J_{\theta_n} = \frac{V_0 Y_n}{2\pi a}\frac{P_n^1(\cos\theta)}{P_n^1(0)} \qquad [1]$$

So in the $n = 1$ mode, $P_1^1(\cos\theta) = \sin\theta$, $P_1^1(0) = 1$.

$$J_{\theta_1} = \frac{V_0 Y_1}{2\pi a}\sin\theta \qquad [2]$$

The average power dissipated in the antenna owing to this current flowing in a conductor of surface resistivity R_s is

$$P_L = 2\int_0^{\pi/2} 2\pi a \sin\theta \frac{|J_\theta|^2 R_s}{2} a\, d\theta$$

$$= \frac{V_0^2|Y_1|^2 R_s}{2\pi}\int_0^{\pi/2} \sin^3\theta\, d\theta = \frac{V_0^2|Y_1|^2 R_s}{3\pi} \quad [3]$$

The average power radiated in the first mode may be written

$$P_R = \frac{V_0^2 G_1}{2} \quad [4]$$

where G_1 is the conductance, or real part of the admittance calculated previously. The ratio of power dissipated to power radiated, considering only this first mode, is then

$$\frac{P_L}{P_R} = \frac{2|Y_1|^2 R_s}{3\pi G_1} \quad [5]$$

This result may be interpreted as a radiation efficiency. For example, at $ka = 1$, $G_1 = 0.0125$ mho, $|Y_1|^2 = 3.12 \times 10^{-4}$ mho^2. Take R_s for copper at 3000 mc per sec which is 0.014 ohm. The ratio of (5) is then 0.742×10^{-4} corresponding to a radiation efficiency of better than 99.99 per cent.

The above calculation is, of course, only approximate since it has not considered total current or total radiation, but in the manner used it certainly gives the correct order of magnitude, and assures us of these two points:

1. The antenna is extremely efficient as a radiating device.
2. The analysis based on perfect conductivity must be exceedingly good.

Problem 11.29(*a*). Calculate losses more accurately at $ka = 1$, considering more terms in the series, and compare with the above approximate result.

Problem 11.29(*b*). Calculate losses approximately in the neighborhood of $ka = 3$.

11.30 Spheroidal Antennas

Stratton and Chu[13] have given not only solutions for spherical antennas but also for prolate spheroidal antennas. Such a solution includes all spheroidal shapes between the sphere just studied and a thin wire (Fig. 11.30*a*).

The assumptions of Stratton and Chu are those used in the spherical

[13] *Journ. Appl. Phys.*, March, 1941.

antenna of the previous articles. Axial symmetry is assumed, and voltage is applied across a very small gap at the center. Results are quite similar in nature but different in magnitude from the results for the sphere. Input admittance may again be expressed as the sum of a large number of input admittances, one for each harmonic mode of oscillation of the antenna. However, for large eccentricities (large ratios of length to diameter) the resonances of each of these modes are very sharp, as contrasted to the broad resonances of the sphere. At a given order of resonance ($n = 1, 3, 5$, etc.) the other modes are correspondingly less important than in the sphere, so that the resonant mode practically determines the antenna characteristics in the neighborhood of resonance.

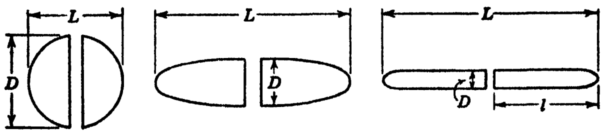


FIG. 11.30. Transition from sphere to thin spheroidal wire dipole.

In the limit of an infinitesimally thin wire, the nth mode becomes resonant slightly below $L = n\lambda/2$. These are true resonances in that the susceptance component of Y_n actually goes through zero and becomes inductive for frequencies above resonance, whereas for the sphere it is always positive (capacitive). At frequencies much higher than resonance, susceptance in the nth mode approaches zero and conductance approaches a small but constant value. This constant value is zero in the limiting case of an infinitesimally small wire, the value found in Prob. 11.28a in the limiting case of a sphere, and something between for medium eccentricities.

These characteristics result in the final curve for input impedance in the neighborhood of the first resonance, as plotted by Stratton and Chu, in their Fig. 2. Again it must be realized that the harmonics cannot be combined exactly until the exact size of gap and exact distribution of applied voltage across the gap is known, yet the form of the curves is accurate and magnitudes are very nearly correct unless the gap is infinitesimal (in which case an infinite input capacitance must be obtained). Curves are given for several values of L/D, of course approaching the results for the spherical antenna as L/D approaches unity.

A study of the curves shows many features associated with past antenna knowledge. The radiation resistance for $L/\lambda = 0.5$ (a half-wave dipole) is found to be about 72 ohms, near the value calculated previously. This varies little for any eccentricity. The condition for zero reactance occurs at something less than a half wavelength for long

thin wires (about 0.49λ or 98 per cent of the antenna length). For fatter wires this condition of zero reactance actually may occur for L greater than a half wavelength (or at a higher frequency than before). In the limit of the sphere there is no place at which input reactance is zero; there is always a capacitive component.

The increase in broadness of the impedance curve is evident for the fatter antennas, thus making available the wide band width required for television antennas.

Finally, Stratton and Chu have plotted the actual current distribution along the antenna for a thin spheroid (large L/D) and found it to vary little from the sinusoidal distribution usually assumed in the conventional methods of calculating antennas.

We will be able to compare some of these results with those obtained by Schelkunoff in the later articles.

11.31 The Biconical Antenna; Equivalent Circuit for Input Impedance

The straightforward solution of Maxwell's equations subject to boundary conditions of the antenna has led to results which are of great interest, especially those results from the study of spheroidal antennas undertaken by Stratton and Chu. However, these results are not easily used for antennas of other more general shapes. We consequently return to a study of the viewpoint developed by Schelkunoff which is particularly rich in physical pictures. The rigorous field analysis in this method is performed on the biconical antenna, consisting of two coaxial cones placed tip to tip with an infinitesimal gap between apices, across which excitation may be applied. The principal and higher order waves on these cones have been discussed with reference to this problem earlier in this chapter (Art. 11.03). The next step is to show that input impedance for the antenna may be calculated exactly from an equivalent circuit in which a transmission line of length l (the cone length) is drawn for the principal wave, and the effect of local waves is obtained from a lumped terminating impedance at the end, $r = l$.

The form of solutions for the outer region (Fig. 11.31a) is exactly that for the spherical wave types developed in Art. 11.25. That is, axial symmetry ($\partial/\partial\phi = 0$) will be assumed and only the TM wave components H_ϕ, E_r, and E_θ will be excited; the axis ($\theta = 0, \pi$) is included in this region, so only the P_n^1 (cos θ) functions are required, and n must be an integer; the region extends outward to infinity, so the second Hankel function will be used for the Bessel function solution. Equation 11.25(19) may then be used directly for the region $r > l$ with $Z_{n+\frac{1}{2}}$ read as $H_{n+\frac{1}{2}}^{(2)}$.

For the region between the cones, $r < l$, there is the principal wave, and to this must be added higher order TM waves similar to those in the space outside the antenna. The TM waves for this region will, however, be somewhat different in form. The Bessel function solution in this region can contain only a $J_{n+\frac{1}{2}}$ term since $N_{n+\frac{1}{2}}$ becomes infinite at $r = 0$. For future purposes note that all field components in these higher order waves then disappear at $r = 0$ since $J_{n+\frac{1}{2}}(0) = 0$. Moreover, a second Legendre function solution is required for this region to

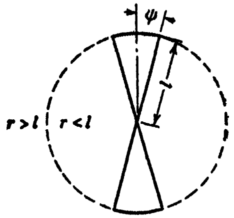


FIG. 11.31a. Biconical antenna.

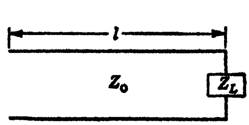


FIG. 11 31b Equivalent circuit of biconical antenna.

account for the two boundary conditions of the cones at $\theta = \psi$ and $\pi - \psi$. This second solution is usually denoted Q_n^1 (cos θ). Its value of infinity on the axis does not trouble us since the axis is excluded from the dielectric region over which the wave solution is to apply by the conducting cones. Thus with the Bessel function read as $J_{n+\frac{1}{2}}$ and an extra associated Legendre function, the TM waves applicable to the region $r < l$ will be similar to Eq. 11.25(19). The order n (probably better written ν) is in general not an integer because of the presence of the cones.[14] This is in fact determined by the boundary conditions $E_\theta = 0$ at $\theta = \psi, \pi - \psi$.

The total current flow in the cones is proportional to H_ϕ.

$$I(r) = 2\pi r \sin \psi H_\phi \big|_{\theta=\psi} \qquad [1]$$

Since H_ϕ in the region of the cones is made up of contributions from the principal and complementary waves, so is I.

$$I(r) = I_0(r) + I'(r) \qquad [2]$$

I_0 denotes the current from fields in the principal wave, I' from the higher order waves. The latter is zero at $r = 0$, since the higher order

[14] It is then possible to use P^m $(-\cos\theta)$ as a second independent solution in place of Q_ν^m $(\cos\theta)$, as is done by Schelkunoff; he also gives equations in terms of the ordinary Legendre functions rather than the associated, since the two are related by the simple derivative.

wave components disappear at the origin. Total current flow into the antenna at $r = 0$ is then only that in the principal wave.

$$I(0) = I_0(0) \qquad [3]$$

Now define a total voltage between the two conical conductors as the integral of E_θ over a surface r = constant.

$$V(r) = -r\int_{\psi}^{\pi-\psi} E_\theta \, d\theta$$

A study of this integral for the higher order wave components of E_θ would reveal that the net integral is zero for all such waves at any radius and has a contribution only for the principal wave. The corresponding situation is readily seen in the higher order waves between parallel planes or rectangular wave guides, where the sinusoids representing transverse electric fields yield just as much negative as positive contribution to the integral of electric field, and so give a net integral of zero.

$$V(r) = V_0(r) \qquad [4]$$

Finally if total current is zero at the end of the antenna with $r = l$, (2) then requires

$$I(l) = 0 \quad \text{or} \quad I_0(l) = -I'(l) \qquad [5]$$

Thus if an equivalent transmission line circuit is drawn to represent the behavior of the principal wave, current in this wave at the end must not be zero as was supposed in drawing the first approximation of an open-circuited line in Art. 11.03, but has the value given by (5). We can assure that this value will be obtained from the equivalent circuit by placing an impedance Z_L across the line at $r = l$ where

$$Z_L = \frac{V_0(l)}{-I'(l)} \qquad [6]$$

Thus the behavior of the principal wave is exactly described by the equivalent circuit of Fig. 11.31*b*, where Z_L is defined by (6). Moreover, since input current in the principal wave is exactly the total input current by (3), and voltage in the principal wave is total voltage everywhere by (4), the input impedance calculated from this principal wave equivalent circuit is the total input impedance. Of course it is not necessary that total current be zero at the end of the antenna in order for the equivalent circuit to be of use. For a finite $I(l)$, $I(l) - I'(l)$ is the current to be accounted for by the lumped impedance.

Although the equivalent circuit has been shown to be exact for calculation of input impedance for any given condition, to show that it is use-

ful requires proof that the impedance Z_L defined by (6) is not a function of the voltage, and that it can be calculated at least approximately for a given antenna system in terms of the dimensions of the antenna. These two requirements are met, as will be seen next.

Problem 11.31(*a*). Consider a biconical antenna with $l = \lambda/4$. From the approximate value of power radiated from a half-wave dipole in Art. 11.09, find the expression for the appropriate value of resistance to use for Z_L to account for this radiated power, if the reactive part of this is neglected. For a thin antenna $\psi = 0.1°$, and a thick antenna $\psi = 5°$, find the values of this resistance. Assuming that the resistance does not vary appreciably with frequency over a small range, plot input impedance of the antenna over a small range about the resonant length $l = \lambda/4$ for the two antennas. What conclusion do you draw on impedance band width of thick antennas versus slender ones?

Problem 11.31(*b*). Assuming that the foreshortening of the thin spheroidal antennas mentioned in Art. 11.30 is of the proper order of magnitude for the conical antenna, what order of reactance component for Z_L do you predict for the above two antennas? Compare with the resistance component.

11.32 Radiation Impedance of Biconical Antenna

In order to calculate Z_L in the equivalent circuit of Fig. 11.31*b*, Schelkunoff has shown two methods. In the first method, the complex Poynting flow of power from an infinitesimal biconical antenna is computed, and this is interpreted in terms of an input impedance. By comparing the result with the expression for input impedance in the equivalent circuit with $Z_0 \to \infty$, Z_L is identified as

$$Z_L = \frac{Z_0^2}{G(kl) + jF(kl)} \qquad [1]$$

where $k = \omega\sqrt{\mu\epsilon} = 2\pi/\lambda$. $G(kl)$ and $F(kl)$ are functions of the electrical length and are plotted in Fig. 11.32*a*.

The second method shown by Schelkunoff follows from a more direct study of the higher order waves. The matching operation of the wave solutions is carried through approximately, in a manner applicable to cones of high characteristic impedance. The steps are as follows.

1. It is assumed that field at a large distance from the antenna is of the same form as that found previously (Art. 11.07) for a dipole antenna. This function of θ is expanded in a series of Legendre polynomials in $\cos\theta$ by the rules used in Art. 11.27.

2. By noting the limiting case of the wave solutions, Eq. 11.25(19), for large values of kr, the unknown coefficient of the nth order term may be evaluated by comparing with the corresponding term in the series from step 1. Coefficients are of course proportional to principal wave current or voltage. Thus the field in the region $r > l$ is in reality

taken as that field found previously from the integration of effects from the assumed sinusoidal distribution of current, but now expanded as a sum of TM waves.

3. Now it is next noted that the TM wave solutions inside the cone region approach exactly the corresponding waves in the space outside as $Z_0 \to \infty$. Thus for matching of E_r across the boundary, the coefficients

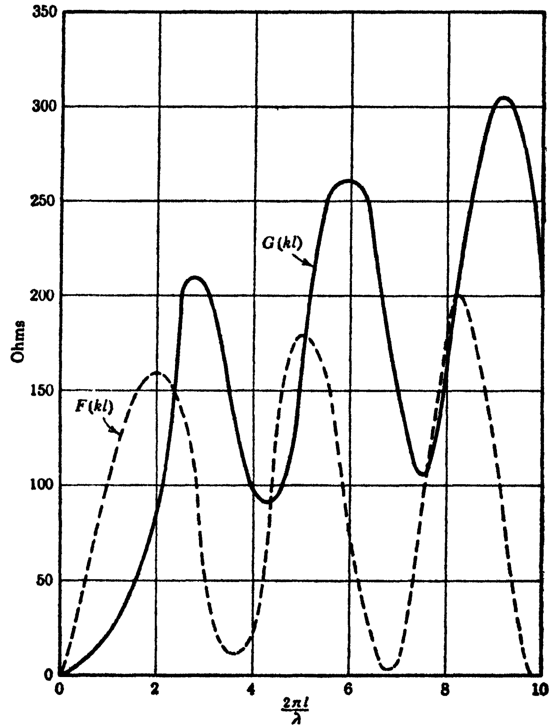


FIG. 11.32a.

of corresponding wave orders inside and outside must be equal in the limit of $Z_0 = \infty$ since there is no E_r component in the principal wave. For large but finite values of Z_0, the coefficients inside may then be taken as equal to those outside to a first approximation. Coefficients of higher order waves inside the antenna region are then obtained in terms of the principal wave current or voltage. Thus the step 3 of the converging step-by-step method sketched in Art. 11.04 is performed, at least approximately.

4. Since coefficients for H_ϕ of the higher order waves are now determined, current in the cones due to these is given by Eq. 11.31(1). Once

this current is determined and written in terms of voltage in the principal wave at $r = l$, Z_L is given by Eq. 11.31(6). The method again gives the same form as in (1), and

$$G(kl) = \sum_{m=0}^{\infty} b_m J^2_{2m+\frac{3}{2}}(kl)$$

$$F(kl) = -\sum_{m=0}^{\infty} b_m J_{2m+\frac{3}{2}}(kl) N_{2m+\frac{3}{2}}(kl)$$

where

$$b_m = \frac{30\pi kl(4m+3)}{(m+1)(2m+1)} \qquad [3]$$

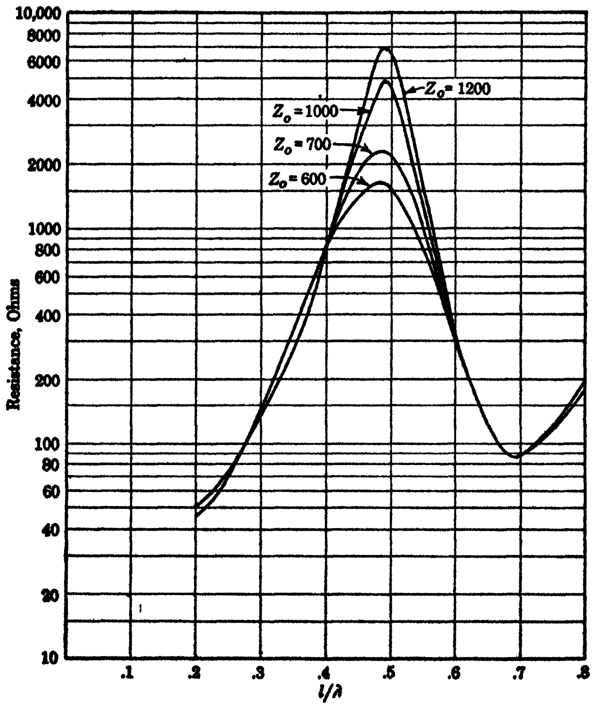


FIG. 11.32*b*. Input resistance of biconical antennas.

Curves showing the resistance and reactance components of input impedance for biconical antennas as a function of length and characteristic impedance are shown in Figs. 11.32*b* and *c*. These are calculated from ordinary transmission line theory for the equivalent circuit Fig. 11.31*b* with Z_L defined by the above. These important practical points follow.

1. Input resistance in the neighborhood of the first resonance is close to the value 73 ohms for a half-wave dipole (Art. 11.09) regardless of the size of the antenna.

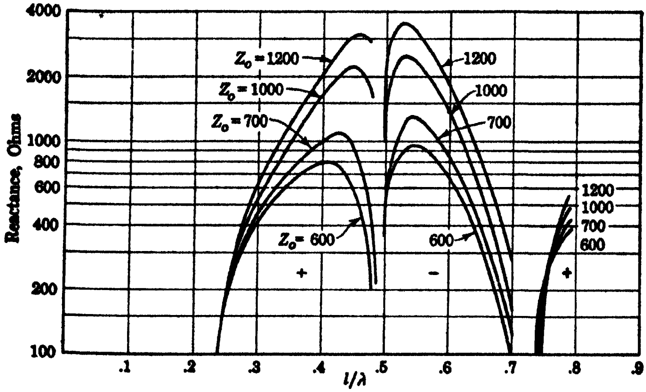


FIG. 11.32c. Input reactance of biconical antenna.

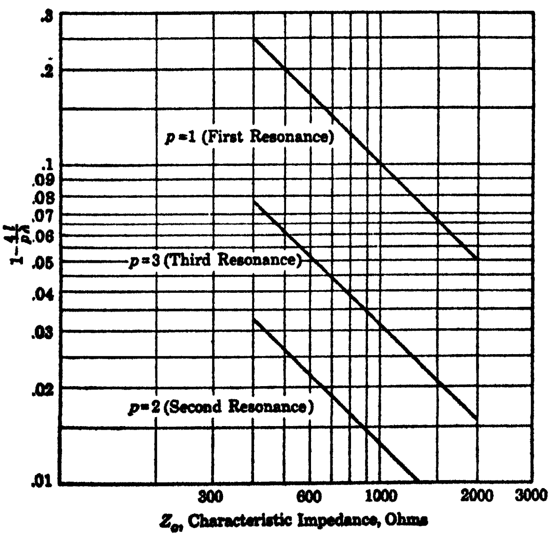


FIG. 11.32d. Per unit foreshortening of biconical antenna.

2. Resonance occurs for the antenna somewhat shorter than the corresponding integral number of half waves, this shortening being greater for the lower characteristic impedances. A curve of shortening versus Z_0 for the first three resonances is given in Fig. 11.32d.

3. Resonance is sharper for high characteristic impedances, again

demonstrating the broad band input impedance properties of the fatter antennas.

4. In the neighborhood of the second resonance (high driving point impedance antennas) input resistance is a definite function of Z_0 as shown by Fig. 11.32*b*.

Problem 11.32. Expand the field of a dipole antenna calculated approximately in Art. 11.07 in a series of spherical transverse magnetic waves.

11.33 Antennas of General Shape

Schelkunoff has extended results of the analysis based on the biconical antennà to antennas of other shape. The method is approximate, but if antennas are not of too great diameter to length ratio, these approximations are easy to accept on a physical basis. It is assumed that the same equivalent circuit applies (Fig. 11.31*b*), but the shape of the antenna is taken into account by considering the antenna as a non-uniform transmission line. For example, if the antenna is cylindrical (Fig. 11.33*a*), the capacity and inductance per unit length may be obtained approximately at any radius by considering the values for a cone that would just pass through this radius. For small ratios of a/r (a = antenna radius, r = distance along antenna from center),

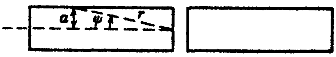


FIG. 11.33*a*. Cylindrical dipole interpreted as a non-uniform transmission line.

$$L \cong \frac{\mu}{\pi} \ln \frac{2}{\psi} = \frac{\mu}{\pi} \ln \frac{2r}{\rho} \qquad C \cong \frac{\pi\epsilon}{\ln \dfrac{2r}{\rho}} \tag{1}$$

Thus L and C, and hence Z_0, are functions of r.

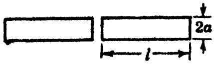
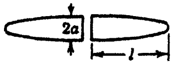
As a first approximation, all previous curves plotted for the biconical antennas may be used, with characteristic impedance taken as an average value over the length defined by

$$Z_{0a} = \frac{1}{l} \int_0^l Z_0(r)\, dr \tag{2}$$

Formulas for the average characteristic impedance for cylindrical, spheroidal, and diamond shape longitudinal-sectioned wires are given in Table 11.33.

The first approximation, using only average characteristic impedance, might be improved by using the same value of Z_L found for the conical antenna, employing non-uniform transmission line theory to find input impedance. The most important correction is probably the correction

TABLE 11.33

Antenna Shape	Thin Cylinders	Thin Spheroids	Diamond Shaped Longitudinal Section
Average characteristic impedance	$120\left(\ln \frac{2l}{a} - 1\right)$	$120 \ln \frac{l}{a}$	$120 \ln \frac{2l}{a}$
Per-unit shortening at first resonance	$\frac{27.08}{Z_{0a}}$	$\frac{5040}{(Z_{0a}+83)^2}$	$\frac{27.08}{\pi Z_{0a}}$
Per-unit shortening at second resonance	$\frac{39.92}{Z_{0a}}$	$\frac{25.68}{Z_{0a}}$	$\frac{30.82}{Z_{0a}}$

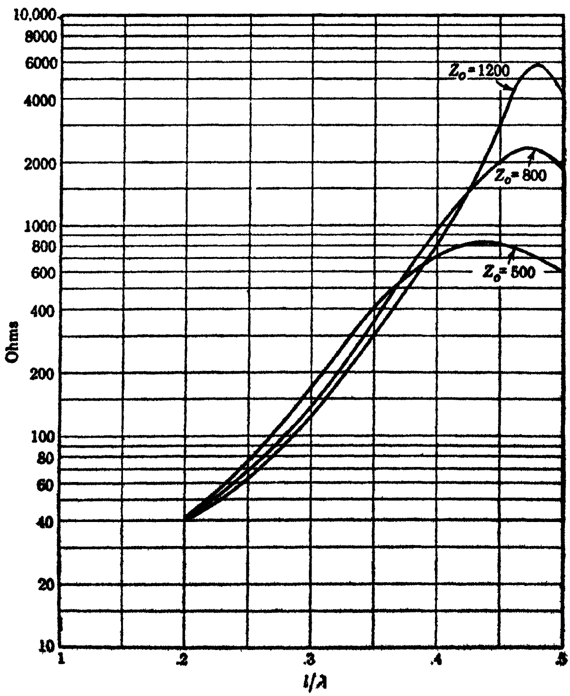


Fig. 11.33b. Input resistance for cylindrical dipole antenna.

to resonant length arising from the non-uniformity of the transmission line. This may be found approximately by other methods. There is found to be a correction to resonant length due to the antenna shape

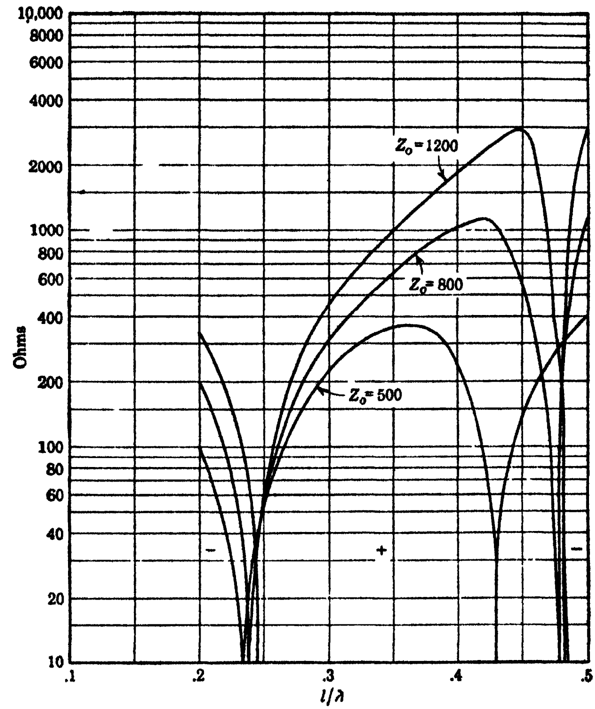


Fig. 11.33c. Input reactance for cylindrical dipole antenna.

which may be either in the same direction or the opposite direction to the correction from the terminating reactance. Approximate formulas for the net length at first and second resonances are also listed in Table 11.33.

Non-uniform transmission line theory applied to the cylindrical antennas gives the input resistance and reactance curves of Fig. 11.33b and c as a better approximation than that from using average characteristic impedance and results for biconical antennas.

11.34 Impedance of Antennas above Earth

Previous results may be applied to antennas above earth if the earth may be assumed perfectly conducting for the first approximation. The image of the antenna in the plane perfectly conducting earth then gives the corresponding free-space antenna configuration to which previous

theory applies. Input impedance of the actual antenna is then half that calculated from the free-space configuration. For example, a vertical cylinder above earth is analyzed as a cylindrical dipole (Fig. 11.34), and resulting input impedance is divided by two.

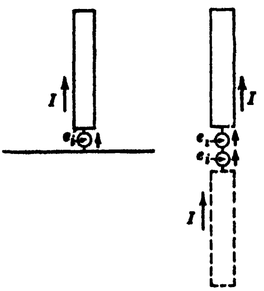


FIG. 11.34. Use of image for determining impedance of antenna above plane conducting earth.

11.35 Current Distribution Along the Antenna

In addition to the final and very practical results from the previous curves, one may ask about total current distribution and its comparison with the principal wave current used in approximate antenna analyses throughout the chapter. From a study of the higher order waves in the region of the antenna, it is found that current due to these reduces to zero as the antenna cross section becomes infinitesimal. This is consistent with the statement made in Art. 11.32 that these waves, in the region of the antenna, approach the distribution of the corresponding wave orders in the space outside the antenna as characteristic impedance approaches infinity, showing that they are unaffected by the infinitesimal conductor. Thus in the limit of infinitesimal cross section, the current distribution is sinusoidal. This of course does not mean that the higher order waves are unimportant in their effects on field distributions.

Schelkunoff has plotted total current distributions for biconical antennas with a characteristic impedance of 1000 ohms. The figures show that the real part of total current is very close to that in the principal wave, although the imaginary parts differ appreciably. The magnitude of total current still compares quite well with that in the principal wave. The deviation will, of course, be more marked as the antenna characteristic impedance decreases.

These results are consistent with the curve of Stratton and Chu which shows current in the thin spheroidal antenna to be very close to a sinusoidal distribution.

Problem 11.35. If current in the higher order waves approaches zero as $Z_0 \to \infty$, the terminating impedance Z_L in the equivalent circuit approaches infinity. It seems that the possibility of accounting for radiation in the equivalent circuit is then excluded. Demonstrate that such reasoning is faulty.

APPENDIX

A. SOME USEFUL REFERENCES

I. Electromagnetic Theory

Schelkunoff, " Electromagnetic Waves," Van Nostrand, 1943.

Stratton, " Electromagnetic Theory," McGraw-Hill, 1941.

Smythe, " Static and Dynamic Electricity," McGraw-Hill, 1939.

Page and Adams, " Electrodynamics," Van Nostrand, 1940.

Skilling, " Fundamentals of Electric Waves," John Wiley, 1942. (Introductory; higher mathematics not required)

II. Mathematical Tools

Dwight, " Tables of Integrals and other Mathematical Data," Macmillan, 1934.

Peirce, " A Short Table of Integrals," Ginn and Co., 1929.

Jahnke and Emde, " Tables of Functions," Dover Publications, 1943.

McLachlan, " Bessel Functions for Engineers," Oxford, 1934.

Byerly, " Harmonic Functions," John Wiley, 1906.

Byerly, " Fourier Series and Spherical Harmonics," Ginn and Co., 1893.

Rothe, Ollendorf, *et al.*, " Theory of Functions as Applied to Engineering Problems," Technology Press MIT, 1933.

Sokolnikoff, J. S., and Sokolnikoff, E. S., " Higher Mathematics for Engineers and Physicists," McGraw-Hill, 1934.

Doherty and Keller, " Mathematics of Modern Engineering," John Wiley, 1936.

Margenau and Murphy, " The Mathematics of Physics and Chemistry," Van Nostrand, 1943.

III. Electric Circuit Theory

Guillemin, " Communication Networks," (Vols. 1 and 2), John Wiley, 1935.

Gardner and Barnes, " Transients in Linear Systems Studied by the Laplace Transformation," John Wiley, 1942.

IV. Mathematical Physics

Slater and Frank, " Introduction to Theoretical Physics," McGraw-Hill, 1933.

Houston, " Principles of Mathematical Physics," McGraw-Hill, 1934.

B. NOMENCLATURE

Listed below are some symbols which we have attempted to use consistently for representation of certain quantities whenever they arise. Symbols which appear in different roles in many different articles, or the transient use of the symbols below for other purposes are not listed here, but the use should be clear from specific definition in the articles where they appear.

		Page where symbol appears
$\bar{a}$	unit vector (subscript denotes direction)	51
$\bar{A}$	vector magnetic potential	73
$\bar{B}$	magnetic flux density	75
c	velocity of light in free space	165
C	capacitance	2
$Ci(x)$	cosine integral	434
$\bar{D}$	electric displacement flux density	52
e	base of natural logarithms $(2.71828 \cdots)$	6
$\bar{E}$	electric field intensity	51
$E(x)$	complete elliptic integral of second kind	228
f	frequency	
G	conductance	10
h	metric coefficient in general coordinates	86
$\bar{H}$	magnetic field intensity	66
$H_\nu^{(1)}(x)$	Hankel function of order ν, first kind	124
$H_\nu^{(2)}(x)$	Hankel function of order ν, second kind	124
$\bar{i}$	current density	70
I	current	2
$I_\nu(x)$	modified Bessel function of first kind, order ν	118
Im	imaginary part of $\cdots$	439
j	$\sqrt{-1}$	6
J	linear current density	210
$J_\nu(x)$	Bessel function of first kind, order ν	116
k	$\omega\sqrt{\mu\epsilon}$	
$K(x)$	complete elliptic integral of first kind	228
$K_\nu(x)$	modified Bessel function of second kind, order ν	127
l	length	
L	self inductance	2
$\bar{L}$	magnetic radiation vector	457
m, n, p	integers	
M	mutual inductance	222
$\bar{n}$	inward unit vector normal to a surface	210
$N_\nu(x)$	Bessel (Neumann's) function of second kind, order ν	117
$\bar{N}$	electric radiation vector	454
$\bar{P}$	Poynting vector	242
$P_m(x)$	ordinary Legendre function of first kind, order m	139
$P_n^{(m)}(x)$	associated Legendre function of first kind, order n, degree m	469
q	charge	50
q	generalized coordinate	86
Q	quality factor of coil or resonant circuit	9
r	radius (coordinate in cylindrical or spherical coordinates)	85

		Page where symbol appears
R	resistance	9
R_s	skin effect surface resistivity	209
Re	real part of $\cdots$	18
S	surface	53
S	susceptance	14
$Si(x)$	sine integral	434
t	time	
T	period	9
u	real part of W, function of a complex variable	102
U	energy	2
v	imaginary part of W, function of a complex variable	102
v	velocity	25
v_p	phase velocity	46
v_g	group velocity	46
V	voltage	11
V	volume	54
W	power	8
W	complex variable	101
Y	admittance	
x, y, z	rectangular coordinates	54
Z	impedance	15
Z	complex variable	101
Z_0	characteristic impedance	27
α	attenuation constant	32
β	phase constant	35
γ	propagation constant, $\alpha + j\beta$	43
Γ	gamma function	128
δ	skin effect depth of penetration	204
Δ	variation of a quantity	45
ϵ	dielectric constant	50
ϵ'	dielectric constant on basis of space as unity, ϵ/ϵ_0	52
ϵ_0	dielectric constant of space	157
ϵ''	loss factor of a dielectric	274
η	intrinsic impedance of a dielectric, $\sqrt{\mu/\epsilon}$	244
θ	polar angle (colatitude) of spherical coordinates	85
λ	wavelength	35
μ	permeability	157
μ'	permeability on basis of space as unity, μ/μ_0	66
μ_0	permeability of space	158
ν	general order in Bessel equation	127
π	$3.14159 \cdots$	
ρ	charge density	53
ρ_s	surface charge density	57
σ	conductivity	172
Σ	a summation	
ϕ	azimuthal angle in cylindrical and spherical coordinates	85
Φ	scalar potential	61
ω	angular frequency, $2\pi f$	5
∇	del (nabla), operator in vector notation	58

INDEX